Discrete-time option pricing with stochastic liquidity

Leippold, Markus ; Schärer, Steven

Abstract: Classical option pricing theories are usually built on the law of one price, neglecting the impact of market liquidity that may contribute to significant bid-ask spreads. Within the framework of conic finance, we develop a stochastic liquidity model, extending the discrete-time constant liquidity model of Madan (2010). With this extension, we can replicate the term and skew structures of bid-ask spreads typically observed in option markets. We show how to implement such a stochastic liquidity model within our framework using multidimensional binomial trees and we calibrate it to call and put options on the SP 500.

DOI: https://doi.org/10.1016/j.jbankfin.2016.11.014

Discrete-time option pricing with stochastic liquidity

Markus Leippold* Steven Schärer†

September 6, 2016

Abstract

Classical option pricing theories are usually built on the law of one price, neglecting the impact of market liquidity that may contribute to significant bid-ask spreads. Within the framework of conic finance, we develop a stochastic liquidity model, extending the discrete-time constant liquidity model of Madan (2010). With this extension, we can replicate the term and skew structures of bid-ask spreads typically observed in option markets. We show how to implement such a stochastic liquidity model within our framework using multidimensional binomial trees and we calibrate it to call and put options on the S&P 500.

JEL Classification: C51; D52; G12; G13
Keywords: Market Liquidity; Bid-Ask Spreads; Option Pricing; Stochastic Liquidity; Conic Finance

*Department of Banking and Finance, University of Zurich, Switzerland.

†Department of Banking and Finance, University of Zurich, Switzerland.

1 Introduction

Classical option pricing theories are usually based on the paradigm of complete and frictionless markets. However, even in financial markets that are considered to be highly competitive, we do observe drops in liquidity, which in times of financial turmoils may be significant and spark concerns among market participants. Liquidity has many different facets. In this paper, we measure liquidity as the spread between bid and ask prices. Illiquid assets are characterized by a high spread. When illiquidity draws a wedge between bid and ask prices, we can no longer rely on the law of one price.

The first attempts to explain bid-ask spreads were made by introducing transaction costs such as commission charges or inventory costs.[1] However, these models often fail to explain the magnitude of the spreads observed in the markets. Especially after the financial crisis of 2008, bid-ask spreads of many assets were persistently high and at a level that cannot be explained by transaction costs alone.[2] A different approach was taken by Madan and Cherny (2010) which is based on theory of conic finance, originating from the work by Cherny and Madan (2009). The basic premise is that the market takes the role of a central counterparty that buys and sells assets from and to investors. The investor buys at the ask price and sells at the bid price. The difference of these prices gives rise to the bid-ask spread observed in financial markets. The central counterparty is viewed as passive in that it does not maximize some utility function, but rather carries out all trades that are acceptable to it.[3]

Madan and Cherny (2010) propose to model market illiquidity by a single market stress level parameter, according to which the market assigns bid and ask prices to assets based on the concept of acceptability indices. This static liquidity model was further extended and taken to the data in Corcuera, Guillaume, Madan, and Schoutens (2012) and Albrecher, Guillaume, and Schoutens (2013). These papers suggests at least two stylized facts for implied liquidity. First, market liquidity implied by real-world data exhibits both a skew and a term structure. This observation is in stark contrast to the assumption of a single liquidity parameter over all maturities and strikes. Second,

[1] See, e.g., Davis, Panas, and Zariphopoulou (1993), Shreve and Soner (1994), Soner, Shreve, and Cvitanić (1995), Cvitanić and Karatzas (1996), Barles and Soner (1998).

[2] See, e.g., Pedersen (2009) for the average bid-ask spreads of large-cap U.S. stocks from June 2006 to June 2009.

[3] Acceptability itself is measured by acceptability indices, which are rooted in the theory of coherent risk measures as developed in Artzner, Delbaen, Eber, and Heath (1999).

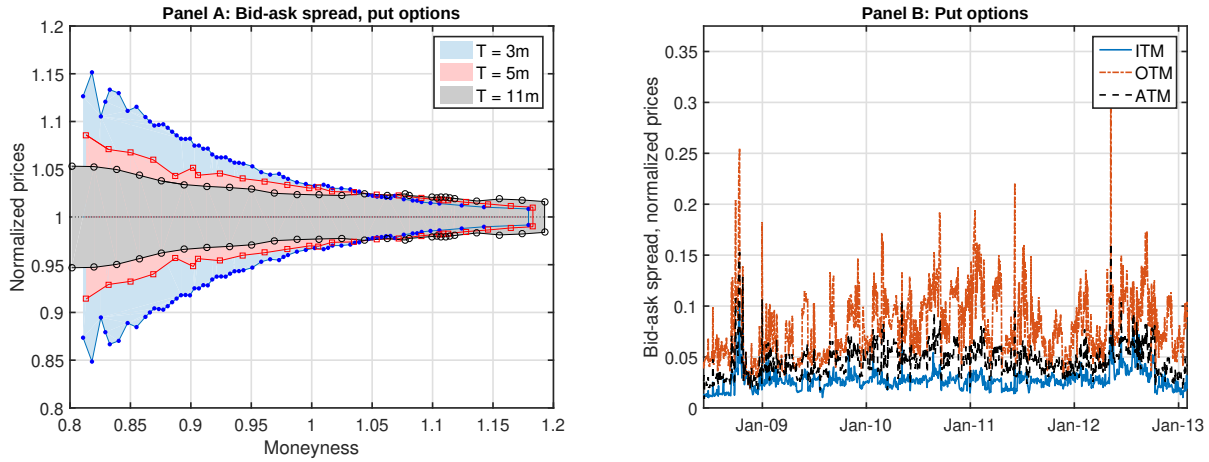

Figure 1. In Panel A, we plot the bid-ask spread for selected maturity slices of the European puts on the S&P 500 on July 20, 2012. We normalize the bid-ask spreads by the mid-prices. Moneyness is expressed as the ratio of strike over forward price. In Panel B, we plot the time-series of bid-ask spreads of in-the-money (ITM, 120% moneyness), out-of-the-money (OTM, 80% moneyness), and at-the-money (ATM, 100% moneyness) puts on the S&P500 with maturity of 5 months on July 20, 2012.

they show that when we calibrate a single market liquidity parameter for the S&P 500 option market, we obtain a time-series of the implied liquidity parameter with a mean-reverting stochastic behavior over time.

These stylized facts are illustrated in Figure 1. In Panel A, we plot the bid-ask spreads in terms of normalized prices of European puts written on the S&P 500 index. Clearly, bid-ask spreads differ across moneyness and time-to-maturity. In Panel B of Figure 1, we plot the historical bid-ask spreads for European puts with a maturity of five months. Clearly, these historical spreads change over time and exhibit some mean-reverting behavior. Hence, the empirical evidence presented in Corcuera, Guillaume, Madan, and Schoutens (2012) and Albrecher, Guillaume, and Schoutens (2013), together with the snapshot of historical bid-ask spreads in Figure 1, provides us with valuable guidance in designing a stochastic liquidity model that may account for the skew and term structure effects of implied liquidity.

We contribute to the steadily growing literature on liquidity modeling for option pricing in two ways. First, by making the setup of the discrete-time constant liquidity model of Madan (2010) more rigorous, we can simplify his results and extend the constant liquidity model to a stochastic liquidity framework. Our Theorem 1 allows us to represent bid and ask prices under stochastic liquidity given by backward recursions as time-consistent and dynamically translation invariant nonlinear expecta-

tions. This result opens the door to introduce stochastic liquidity in the conic finance framework. As an illustration, we apply a specific stochastic liquidity model using multidimensional binomial trees to the S&P 500 index option market. We show that this extension improves the fit of the term and skew structures in bid-ask spreads observed in markets. To the best of our knowledge, this is the first model to treat liquidity as a separate process that can be applied to the pricing of bid-ask spreads of derivatives. Compared to other approaches, our model is also suitable for deriving the bid and ask prices of path-dependent options such as Asian and Barrier options.

There have been various other endeavors on how to introduce dynamic bid-ask spreads in option pricing based on the model of Madan and Cherny (2010). An obvious way to do so is to model the bid and ask price as two separate stochastic processes, as suggested in Madan and Schoutens (2014). However, it can be considered a drawback that for payoffs which are not comonotone with a long or short stock position, this approach only gives lower and upper bounds for bid and ask prices. Another avenue is to follow the literature of dynamic risk measures. Mirroring the steps of the static one-step model, Bielecki, Cialenco, Iyigunler, and Rodriguez (2013) define dynamic acceptability indices with the help of dynamic coherent risk measures as discussed in, e.g., Riedel (2004) and Artzner, Delbaen, Eber, Heath, and Ku (2007). The disadvantage of using dynamic coherent risk measures is that they are not as tractable and intuitive compared to the static setup. It is furthermore not clear how a stochastic liquidity component could be incorporated. Biagini and Bion-Nadal (2014) tackle the issue in a similar way and arrive at a continuous-time version, while Bielecki, Cialenco, and Chen (2015) make use of Backward Stochastic Difference Equations (BSΔEs) and Rosazza Gianin and Sgarra (2013) derive dynamic risk measures from g-expectations.

Other than the approaches described above which are all based on or inspired by Conic Finance, there is a large body of literature that explores liquidity and bid-ask spreads in option markets.[4] One way to derive bid and ask prices of derivatives is by considering the replication costs induced by an illiquid underlying. A popular model in this direction was conceived by Çetin, Jarrow, and Protter (2004) who propose to model illiquidity by assuming that prices of underlyings are provided by a stochastic supply curve, that is not impacted by the actions of buyers and sellers. The resulting

[4]See, e.g., George and Longstaff (1993), Engle and Neri (2010), Chou, Chung, Hsiao, and Wang (2011), Chan and Chung (2012), Bongaerts, De Jong, and Driessen (2011), Christoffersen, Goyenko, Jacobs, and Karoui (2015), and Feng, Hung, and Wang (2014), to name a few.

bid and ask prices are then dependent on the trade size which differs from our assumption of a trade-invariant bid-ask spread. They find that, in discrete time, hedging derivatives by trading the illiquid underlying incurs liquidity costs. However, results in Çetin, Jarrow, Protter, and Warachka (2006) indicate that this approach can only partially explain bid-ask spreads of derivatives observed in the market. In a separate study, Chou, Chung, Hsiao, and Wang (2011) also conclude that it is not sufficient to only consider the underlying's liquidity, but also an option's own liquidity must be taken into account. In contrast, our model does not specifically differentiate between underlying and option liquidity and indeed does not use replicating strategies to derive option prices. Hence, we assume that all information regarding liquidity is contained in the bid and ask prices of the option market.[5]

The rest of the paper is structured as follows. In Section 2, we review the one-period framework of Madan and Cherny (2010). Section 3 introduces the multi-period model with stochastic liquidity. In Section 4, we bring our model to the data and show that the stochastic liquidity model helps to explain the skew and term structure typically observed in options' bid ask spreads. Finally, Section 5 concludes. All proofs are delegated to the appendix.

2 One-step static liquidity model

We start with a brief description of the one-step liquidity model presented in Madan and Cherny (2010), since it builds the basis of our stochastic liquidity model presented in the subsequent section. To this end, we fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $L^\infty := L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ the space of all essentially bounded and $\mathbb{R}$-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$.[6] By $\mathbb{P}$, we denote the reference probability measure that we assume to be a risk-neutral measure. In a complete market, this probability measure is unique and the price of an asset is given by the $\mathbb{P}$-expectation of its future discounted cash flows, say $Q \in L^\infty$. When market incompleteness drives a wedge between bid and ask prices, we can interpret the bid (ask) price as being caused by an overweighting (underweighting)

[5] In our approach, we cannot differentiate between how much of the illiquidity reflected in the options' bid-ask spreads are due to the illiquidity of the underlying market and how much is due to the option market itself.

[6] This choice is for simplicity only and the results can be generalized to L^p spaces with $1 \leq p < +\infty$. The discrete-time extension will only consider finite spaces, where L^p spaces are equivalent anyway.

of losses and an underweighting (overweighting) of gains relative to the measure $\mathbb{P}$. Hence, we can model this weighting scheme as a distortion to the reference probability measure. For this purpose, we define a distortion function as follows.

Definition 1 (Distortion function)**.** *A function $\psi : [0,1] \longrightarrow [0,1]$ is a distortion function if and only if it is monotone, $\psi(0) = 0$ and $\psi(1) = 1$.*

The distorted probability measure $\psi \circ \mathbb{P}$ is no longer a probability measure in general. It is, however, still a finite monotone set function that is submodular, if the distortion function is concave. It is therefore possible to define a risk measure based on distorted probabilities using Choquet integrals.[7]

Definition 2 (Distortion risk measure)**.** *Let ψ be a concave distortion function and Q a future discounted cash flow. The function $\varrho^{\psi} : L^{\infty} \longrightarrow \mathbb{R}$ given by*

$$\varrho^{\psi}(Q) := \int_{-\infty}^{0} \psi(\mathbb{P}(Q \leq x))dx - \int_{0}^{\infty} (1 - \psi(\mathbb{P}(Q \leq x)))dx, \quad \forall Q \in L^{\infty}, \tag{1}$$

is called a distortion risk measure induced by ψ.

From the properties of the Choquet integral and because $\psi \circ \mathbb{P}$ is submodular, the function ϱ^{ψ} defined by equation (1) is monotone, positively homogeneous, translation invariant, and subadditive. Hence, it is a coherent risk measure. By inverting the sign of ϱ^{ψ} we obtain what is called a distorted expectation, corresponding to the intuition of weighting losses and gains differently compared to $\mathbb{P}$. In particular, we call the function $\mathbb{E}^{\psi}[\cdot] : L^{\infty} \longrightarrow \mathbb{R}$ given by

$$\mathbb{E}^{\psi}[Q] := -\varrho^{\psi}(Q), \quad \forall Q \in L^{\infty}, \tag{2}$$

the distorted expectation induced by ψ.

Just as coherent risk measures, $\mathbb{E}^{\psi}[\cdot]$ is nonlinear, i.e., in general $\mathbb{E}^{\psi}[Q^1+Q^2] \neq \mathbb{E}^{\psi}[Q^1]+\mathbb{E}^{\psi}[Q^2]$ for $Q^1, Q^2 \in L^{\infty}$. Nevertheless, $\mathbb{E}^{\psi}[\cdot]$ shares many other properties with the usual expectation operator such as monotonicity, positive homogeneity, and translation invariance. The concept of nonlinear

[7] See Choquet (1953). For more details on the properties of Choquet integrals, see the standard book of Denneberg (1994).

expectations, albeit in a multi-period setting, will be the cornerstone of our stochastic liquidity model. In particular, from Definition 2 it follows that for $Q \in L^\infty$ and ψ a concave distortion function,

$$\mathbb{E}^\psi[Q] \leq \mathbb{E}[Q] \leq -\mathbb{E}^\psi[-Q], \tag{3}$$

where $\mathbb{E}[\cdot]$ denotes the expectation with respect to the reference pricing measure $\mathbb{P}$. Hence, distorted expectations provide an intuitive basis for the modeling of bid-ask spreads.

To measure the degree of distortion applied to the reference probability measure $\mathbb{P}$, the central counterparty is assumed to have not only one concave distortion function with which it evaluates potential trades, but a whole family $\Psi = (\psi^z)_{z\geq 0}$. This family of concave distortion functions is pointwise increasing in z in that $\psi^{z_1}(\cdot) \leq \psi^{z_2}(\cdot)$ if and only if $z_1 \leq z_2$. Additionally, ψ^0 is assumed to be the identity function. For a given distortion function $\psi^\gamma \in \Psi$ under which cash flows are evaluated, we can interpret γ as the market liquidity level. The more illiquid the market becomes (i.e., the higher γ is), the more distorted the reference probability measure becomes. A liquidity level of zero implies that no distortion at all is applied, which corresponds to perfect liquidity and hence to the complete market case, in which the law of one price holds.[8]

A discounted cash flow $Q \in L^\infty$ is deemed acceptable at γ if and only if $\mathbb{E}^{\psi^\gamma}[Q] \geq 0$. The market is assumed to competitively execute only acceptable trades. Therefore, the bid price $b^{\Psi,\gamma}(Q)$ of a discounted cash flow $Q \in L^\infty$ is the highest price the market is willing to pay for the net position to be acceptable according to the market liquidity level, i.e.,

$$b^{\Psi,\gamma}(Q) := \sup\{b \in \mathbb{R} \,|\, \mathbb{E}^{\psi^\gamma}[Q - b] \geq 0\} = \mathbb{E}^{\psi^\gamma}[Q], \tag{4}$$

where the last equality follows from the translation invariance of coherent risk measures. A similar argument can be made for the ask price and leads to the following definition of bid and ask prices.[9]

Definition 3 (Single-period bid-ask prices)**.** *Let $\Psi = (\psi^z)_{z\geq 0}$ be a pointwise increasing family of concave distortion functions and $\gamma > 0$ the market liquidity level. Then, the ask price of a discounted*

[8]The liquidity measure in our model is not directly defined by observable variables in the market such as trading volume or the bid-ask spreads. Instead, it is inferred from a comparison of market and model implied bid-ask spreads.

[9]This approach differs significantly from deriving bid and ask prices of derivatives based on a replicating trading strategy as in, e.g., Çetin, Jarrow, and Protter (2004).

cash flow $Q \in L^\infty$ is given by

$$a^{\Psi,\gamma}(Q) := -\mathbb{E}^{\psi^\gamma}[-Q] = \int_{-\infty}^{0} (\psi^\gamma(\mathbb{P}(Q > x)) - 1)\, dx + \int_{0}^{\infty} \psi^\gamma(\mathbb{P}(Q > x))dx, \tag{5}$$

and its bid price is

$$b^{\Psi,\gamma}(Q) := \mathbb{E}^{\psi^\gamma}[Q] = -\int_{-\infty}^{0} \psi^\gamma(\mathbb{P}(Q \leq x))dx + \int_{0}^{\infty} (1 - \psi^\gamma(\mathbb{P}(Q \leq x)))\, dx. \tag{6}$$

Without any further assumptions, the bid price is always less than or equal to the ask price. Bid and ask price also envelop the undistorted price, i.e.,

$$b^{\Psi,\gamma}(Q) \leq \mathbb{E}[Q] \leq a^{\Psi,\gamma}(Q) \quad \forall Q \in L^\infty. \tag{7}$$

Madan and Cherny (2010) derive these bid and ask prices from the theory of acceptability indices, which are functions $\alpha : L^\infty \to [0, \infty]$. In particular, they call a net cash flow, or trade, $\widetilde{Q} \in L^\infty$ acceptable at a certain market liquidity level γ if and only if $\alpha(\widetilde{Q}) \geq \gamma$.[10] Cherny and Madan (2009) show that acceptability indices can be represented by a family of coherent risk measures $(\varrho^z)_{z\geq 0}$ that are continuous from above and pointwise increasing in z. For a pointwise increasing set of concave distortion functions $\Psi = (\psi^z)_{z\geq 0}$, such a family is given by the corresponding distortion risk measures $\Phi = (\varrho^{\psi^z})_{z\geq 0}$. To see that Φ is pointwise increasing in z, note that distortion risk measures retain the ordering of the associated distortion function. Furthermore, all distortion risk measures are continuous from above.[11] Hence there exists an acceptability index which corresponds to Φ.

For our model, we assume the conditional value-at-risk (CVaR) as distortion measure.[12] According to, e.g., Föllmer and Schied (2011), the CVaR at level $\alpha \in (0, 1]$ can be expressed as a distortion

[10] Equivalently, $a^{\Psi,\gamma}(Q) = \sup_{\widetilde{P}\in\mathcal{D}} E^{\widetilde{P}}[Q]$, where $\mathcal{D}$ is the convex set of risk measures that are equivalent to $\mathbb{P}$ and are determined by α and γ.

[11] See, e.g., Föllmer and Schied (2011).

[12] Other measures that have been applied in the literature are the Wang distortion function (Wang (2000)), the MinMaxVar-distortion function introduced in Cherny and Madan (2009), or the EssSupExp-distortion function proposed by Bannör and Scherer (2014).

risk measure

$$\mathrm{CVaR}_\alpha(Q) = \varrho^{\widetilde{\psi}^\alpha}(Q) = \int_{-\infty}^{0} \widetilde{\psi}^\alpha(\mathbb{P}(Q \le x))dx - \int_0^\infty (1 - \widetilde{\psi}^\alpha(\mathbb{P}(Q \le x)))dx, \quad \forall Q \in L^\infty, \tag{8}$$

induced by the concave distortion function

$$\widetilde{\psi}^\alpha(u) := \min\left\{\frac{u}{\alpha}, 1\right\}, \quad \forall u \in [0,1]. \tag{9}$$

The distortion functions in our setup have to depend on a parameter taking values in $[0,\infty)$. However, the quantile parameter α is defined on $(0,1]$. We therefore first use the change of variables $x \mapsto 1-x$ to map $(0,1]$ to $[0,1)$ and then apply a sigmoid function, e.g.,

$$\varphi(x) := \frac{x}{\sqrt{1+x^2}}, \quad \forall x \in \mathbb{R}, \tag{10}$$

which bijectively maps values from $[0,\infty)$ to $[0,1)$. With that we get, for $z \ge 0$ and $\alpha = 1 - \varphi(z) \in (0,1]$,

$$\mathrm{CVaR}_\alpha(Q) = \varrho^{\psi_{\mathrm{CVaR}}^{\varphi^{-1}(1-\alpha)}}(Q) = \varrho^{\psi_{\mathrm{CVaR}}^{z}}(Q) = \mathrm{CVaR}_{1-\varphi(z)}(Q), \quad \forall Q \in L^\infty, \tag{11}$$

for the modified concave distortion function

$$\psi_{\mathrm{CVaR}}^z(u) := \min\left\{\frac{u}{1-\varphi(z)}, 1\right\}, \quad \forall u \in [0,1]. \tag{12}$$

As required, the family $\Gamma^{\mathrm{CVaR}} = (\psi_{\mathrm{CVaR}}^z)_{z\ge 0}$ is pointwise increasing in z and ψ_{CVaR}^0 is the identity function.

3 Discrete-time stochastic liquidity model

The model presented in the last section only considers a single time step. Hence, it is of limited practical value. In this section, we extend the model of Madan (2010) to treat market liquidity as a stochastic process instead of a constant to account for the stylized facts of bid and ask spreads. As in the static model, it is our goal to have bid and ask prices that are represented by nonlinear expectations. Due to the multi-period setup, we additionally require that they behave consistently

over time. These nonlinear expectations can also be used to define dynamic risk measures as we show in Remark A.1.

The proof of the main theorem, which we present in Appendix A, relies on the theory of BSΔEs of Cohen and Elliott (2010b), the discrete-time analogue of Backward Stochastic Differential Equations (BSDEs) as developed by El Karoui, Peng, and Quenez (1997). Similar to how Coquet, Hu, Mémin, and Peng (2002) linked g-expectations to solutions of BSDEs, one can show that certain types of nonlinear expectations are solutions to BSΔEs. As in the static setup, these will be used to define the bid and ask prices of discounted cash flows.

Before we present the main result, we introduce some additional notation. We denote by $T > 0$ maturity and assume we have time points $0 = t_0 < \cdots < t_K = T$ for $K > 0$. By $\mathcal{T}_i^j := \bigcup_{i \leq l \leq j} \{t_l\}$ we denote the set of time points from t_i to t_j, where $0 \leq i, j \leq K$. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}_0^K}, \mathbb{P})$ be a finite filtered probability space satisfying the usual conditions. The underlying price $S = (S_t)_{t \in \mathcal{T}_0^K}$ is modeled as a positive but otherwise general finite state price process on the probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}_0^K}, \mathbb{P})$. As in the static setting, $\mathbb{P}$ denotes the reference probability measure that is assumed to be risk-neutral.

3.1 Time-consistent nonlinear expectations

In the spirit of Peng (2007), dynamic nonlinear expectations are defined as follows.

Definition 4 (Time-consistent nonlinear expectation)**.** *A family of functions $\mathcal{E}(\cdot \,|\, \mathcal{F}_t) : L^1(\mathcal{F}_T) \longrightarrow L^1(\mathcal{F}_t)$ for $t \in \mathcal{T}_0^K$ is a time-consistent nonlinear expectation if it satisfies, for all $s \in \mathcal{T}_0^K$ and $Q, Q^1, Q^2 \in L^1(\mathcal{F}_T)$,*

(i) monotonicity, i.e., if $Q^1 \leq Q^2$ $\mathbb{P}$-a.s., then

$$\mathcal{E}(Q^1 \,|\, \mathcal{F}_t) \leq \mathcal{E}(Q^2 \,|\, \mathcal{F}_t) \quad \mathbb{P}\text{-a.s}$$

Additionally, for $Q^1 \leq Q^2$ $\mathbb{P}$-a.s., equality holds if and only if $\mathbb{P}$-a.s. $Q^1 = Q^2$.

(ii) adaptability, i.e., $\mathcal{E}(Q \,|\, \mathcal{F}_t) = Q$ if Q is $\mathcal{F}_t$-measurable.

(iii) dynamic consistency, i.e., $\mathcal{E}(\mathcal{E}(Q \,|\, \mathcal{F}_t) \,|\, \mathcal{F}_s) = \mathcal{E}(Q \,|\, \mathcal{F}_s)$, $\mathbb{P}$-a.s. $\forall s \leq t$.

(iv) relevance, i.e., $\mathbb{1}_A \mathcal{E}(Q \,|\, \mathcal{F}_t) = \mathcal{E}(\mathbb{1}_A Q \,|\, \mathcal{F}_t)$, $\mathbb{P}$*-a.s.* $\forall A \in \mathcal{F}_t$.

The monotonicity property states that of two different payoffs, the one that is $\mathbb{P}$-a.s. smaller also has a smaller expectation under $\mathcal{E}(\cdot \,|\, \mathcal{F}_t)$ for all $t \in \mathcal{T}_0^K$. Adaptability is assumed due to our multi-period setup. Dynamic consistency is what is known as the tower property for usual conditional expectations. For nonlinear expectations, we have to explicitly make this assumption to ensure that different time steps are linked consistently. Finally, "relevance" means that at time t, the investor knows whether the underlying's path is in $A \in \mathcal{F}_t$. If this is the case, the nonlinear expectation of $\mathbb{1}_A Q$ is the same as the one of Q. Otherwise, it is zero.

Definition 5 (Dynamic translation invariance)**.** *A time-consistent nonlinear expectation* $(\mathcal{E}(\cdot \,|\, \mathcal{F}_t))_{t \in \mathcal{T}_0^K}$ *is dynamically translation invariant if and only if for all* $t \in \mathcal{T}_0^K$*,*

$$\mathcal{E}(Q + q \,|\, \mathcal{F}_t) = \mathcal{E}(Q \,|\, \mathcal{F}_t) + q, \quad \forall Q \in L^1(\mathcal{F}_T), q \in L^1(\mathcal{F}_t). \tag{13}$$

Dynamic translation invariance is not generally required of nonlinear expectations. Nevertheless, we want to ensure that, e.g., a portfolio consisting of one derivative and some cash has the same bid price as adding the cash to the bid price of the derivative on its own.

3.2 Distorted conditional expectation

Instead of considering only a constant market liquidity level γ as in Madan (2010), we introduce a stochastic process to model market liquidity through time and states.

Definition 6 (Market liquidity process)**.** *The time- and state-dependent process* $\Gamma = (\gamma_t)_{t \in \mathcal{T}_0^K}$*, for* $\gamma_t \in L^1(\mathcal{F}_t)$ *and* $\gamma_t \geq 0$ $\mathbb{P}$*-a.s. for all* $t \in \mathcal{T}_0^K$ *is called market liquidity process.*

Before, we used the liquidity level $\gamma \geq 0$ to determine the degree of distortion applied to $\mathbb{P}$ by choosing ψ^γ from a family of pointwise increasing concave distortion functions Ψ. We will continue in this spirit, but since the liquidity level at any time is now a random variable, we introduce a state-dependent distortion function.

Definition 7 (Concave state-dependent distortion function)**.** *A function* $\psi : \Omega \times [0,1] \longrightarrow [0,1]$ *is*

called a concave state-dependent distortion function if and only if for all $\omega \in \Omega$, $\psi(\omega, \cdot)$ *is a concave distortion function.*

Slightly abusing notation, we will denote by $\Psi := (\psi^z)_{z \geq 0}$ the usual family of concave distortion functions that are pointwise increasing in z and define, for a random variable γ,

$$\psi^\gamma(\omega, u) := \psi^{\gamma(\omega)}(u) \quad \forall \omega \in \Omega, u \in [0,1]. \tag{14}$$

This gives a concave state-dependent distortion function as defined above. As such, we will continue working with the same family of concave distortion functions as in the previous section. The notion of distorted expectations is also extended to state-dependent distortion functions. Furthermore, they are now conditional on the filtration.

Definition 8 (Distorted conditional expectation)**.** *Let* ψ *be a concave state-dependent distortion function. The family of functions with elements* $\mathbb{E}_t^\psi[\cdot] : L^1(\mathcal{F}_T) \longrightarrow L^1(\mathcal{F}_t)$ *for* $t \in \mathcal{T}_0^K$*, defined* $\forall Q \in L^1(\mathcal{F}_T)$ *and* $\forall \omega \in \Omega$ *as*

$$\mathbb{E}_t^\psi[Q](\omega) := -\int_{-\infty}^0 \psi(\omega, \mathbb{P}_t(Q \leq x)(\omega))dx + \int_0^\infty 1 - \psi(\omega, \mathbb{P}_t(Q \leq x)(\omega))dx \tag{15}$$

is called distorted conditional expectation.

For brevity, we denote the conditional probability by[13]

$$\mathbb{P}_t(A) := \mathbb{P}(A \,|\, \mathcal{F}_t) = \mathbb{E}[\mathbb{1}_A \,|\, \mathcal{F}_t], \quad \forall A \in \mathcal{F}_T.$$

Since $\mathbb{P}_t(\cdot)(\omega)$ is a probability measure for any state $\omega \in \Omega$ and $t \in \mathcal{T}_0^K$, $\mathbb{E}_t^\psi[\cdot](\omega)$ is a distorted expectation as in the static framework. As in the static case it holds that

$$\mathbb{E}_t^\psi[Q] \leq \mathbb{E}[Q \,|\, \mathcal{F}_t] \leq -\mathbb{E}_t^\psi[-Q], \quad \mathbb{P}\text{-a.s.}, \tag{16}$$

for all $Q \in L^1(\mathcal{F}_T)$, $t \in \mathcal{T}_0^K$, and all concave state-dependent distortion functions ψ.

[13]Note that we assume enough regularity on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathcal{T}_0^K}, \mathbb{P})$ that a regular conditional distribution can be constructed.

3.3 Time-consistent bid and ask prices

To define the bid and ask prices in our multiperiod setting, we borrow from the one-step static liquidity model of Madan and Cherny (2010). We start at the final maturity T and recursively apply the conditional distorted expectation. The stochastic liquidity component thereby determines the re-weighting of the reference probability measure depending on the current state and time step. Working backwards, we arrive at time t_0 and obtain today's bid and ask prices.

Definition 9 (Multi-period bid-ask prices)**.** *For $\Psi := (\psi^z)_{z\geq 0}$ a family of concave distortion functions that are pointwise increasing in z, market liquidity process $\Gamma = (\gamma_t)_{t\in\mathcal{T}_0^K}$ and $t_k \in \mathcal{T}_0^K$, the bid price of a future discounted cash flow $Q \in L^1(\mathcal{F}_T)$ at time t_k is defined as*

$$b_{t_k}^{\Psi,\Gamma}(Q) := \mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[\mathbb{E}_{t_{k+1}}^{\psi^{\gamma_{t_{k+1}}}}[\cdots \mathbb{E}_{t_{K-1}}^{\psi^{\gamma_{t_{K-1}}}}[Q]]]. \tag{17}$$

Its ask price at time t_k is given by

$$a_{t_k}^{\Psi,\Gamma}(Q) := -\mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[\mathbb{E}_{t_{k+1}}^{\psi^{\gamma_{t_{k+1}}}}[\cdots \mathbb{E}_{t_{K-1}}^{\psi^{\gamma_{t_{K-1}}}}[-Q]]]. \tag{18}$$

The distorted conditional expectation does not satisfy the tower property. Hence, it is not a time-consistent nonlinear expectation. Therefore, $b_t^{\Psi,\Gamma}(Q) \neq \mathbb{E}_t^{\psi^\gamma}[Q]$ in general. However, it is still adapted, monotone and dynamically translation invariant, as also Lemma A.2 shows. By accounting for the dynamic translation invariance of the driver in the BSΔE, we are able to simplify the result in Madan (2010) and prove that the bid and ask prices from Definition 9 are time-consistent and dynamically translation invariant nonlinear expectations.[14] At the same time, we correct an error in the proof of Madan (2010), see Remark A.2.

Theorem 1. *Let $\Psi := (\psi^z)_{z\geq 0}$ a family of concave distortion functions that are pointwise increasing in z and $\Gamma = (\gamma_t)_{t\in\mathcal{T}_0^K}$ a market liquidity process. Then, the bid and ask prices $(b_t^{\Psi,\Gamma})_{t\in\mathcal{T}_0^K}$ and $(a_t^{\Psi,\Gamma})_{t\in\mathcal{T}_0^K}$ are time-consistent and dynamically translation invariant nonlinear expectations.*

[14] Time-consistency of bid and ask prices prevents round-trip arbitrage opportunities. In particular, buying Q at time t_0 has to cost the same as buying a newly introduced asset that pays the ask price of Q at time t_1. This payoff could then be used to buy the asset at time t_1, after which the two approaches are equivalent.

3.3.1 Constant liquidity model

As a first application of Theorem 1, we can consider a constant liquidity model. In particular, by fixing $\gamma > 0$ to a constant, we obtain the model suggested by Madan (2010) albeit with simplified expressions for the bid and ask price. Assuming a recombining binomial tree model for the stock price process $(S_{t_k})_{k=0}^{K}$ with a constant time step size $h := \frac{T}{K}$, and $k = 1, ..., K$,

$$S_{t_k} := S_0 u^{\sum_{i=1}^{k} \xi_i} d^{k - \sum_{i=1}^{k} \xi_i}, \tag{19}$$

where $S_0 > 0$, $u \geq 1$, $d = 1/u$ and $(\xi_i)_{i=1}^{K}$ are iid random variables with values in $\{0, 1\}$, for which we set the up and down probabilities $p_u := \mathbb{P}(\xi_i = 1)$, $p_d := \mathbb{P}(\xi_i = 0)$. We also have a family of concave distortion functions Ψ that are pointwise increasing. For option payoffs that are path-dependent, it is necessary to go through the tree backward recursively to calculate the bid and ask prices at time zero. For other payoffs such as European vanilla contracts, it is possible to derive closed-form expressions. Given Theorem 1, we obtain the following analytical formulas for the bid and ask prices of European claims.

Corollary 1. *Let the future discounted cash flow be given by a function H such that $Q = H(T, S_T)$. Further assume that H is non-negative and monotonically increasing with S_T (e.g., a European call option). Then,*

$$a_0^{\Psi,\gamma}(Q) = \sum_{i=0}^{K} \binom{K}{i} \psi^{\gamma}(p_u)^i (1 - \psi^{\gamma}(p_u))^{K-i} H(T, S_0 u^i d^{K-i}) \tag{20}$$

and

$$b_0^{\Psi,\gamma}(Q) = \sum_{i=0}^{K} \binom{K}{i} (1 - \psi^{\gamma}(p_d))^i \psi^{\gamma}(p_d)^{K-i} H(T, S_0 u^i d^{K-i}). \tag{21}$$

Similarly, for a non-negative but monotonically decreasing payoff (e.g., a European put option), we have

$$a_0^{\Psi,\gamma}(Q) = \sum_{i=0}^{K} \binom{K}{i} \psi^{\gamma}(p_d)^i (1 - \psi^{\gamma}(p_d))^{K-i} H(T, S_0 u^{K-i} d^i) \tag{22}$$

and

$$b_0^{\Psi,\gamma}(Q) = \sum_{i=0}^{K} \binom{K}{i} (1 - \psi^{\gamma}(p_u))^i \psi^{\gamma}(p_u)^{K-i} H(T, S_0 u^{K-i} d^i). \tag{23}$$

The derivation follows the same steps as described in the next section.

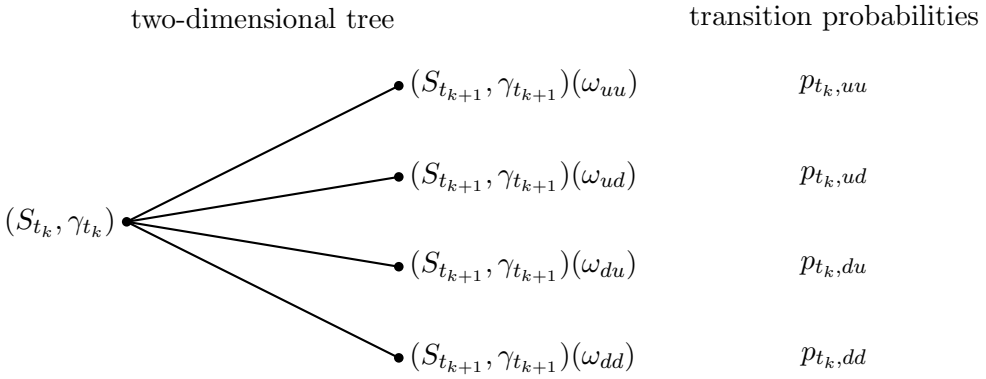

Figure 2. A single node of the two-dimensional binomial tree for (S, γ) with the corresponding transition probabilities.

3.3.2 Stochastic liquidity model

To generalize the previous model to a stochastic liquidity model, we proceed as follows. We describe the dynamics of S and γ by a two-dimensional recombining binomial tree $(S_t, \gamma_t, p_t, q_t)_{t\in\mathcal{T}_0^K}$ as illustrated in Figure 2. For any node (S_{t_k}, γ_{t_k}), there are four connected nodes $(S_{t_{k+1}}, \gamma_{t_{k+1}})(\omega_{uu})$, $(S_{t_{k+1}}, \gamma_{t_{k+1}})(\omega_{ud})$, $(S_{t_{k+1}}, \gamma_{t_{k+1}})(\omega_{du})$ and $(S_{t_{k+1}}, \gamma_{t_{k+1}})(\omega_{dd})$ with corresponding transition probabilities $p_{t_k,uu}$, $p_{t_k,ud}$, $p_{t_k,du}$ and $p_{t_k,dd}$. State ω_{ud} stands for an up-move of the underlying S and a down-move of the liquidity process γ. The other states and transition probabilities are denoted using the same notation.

We are now interested in calculating the bid and ask price of a future discounted cash flow $Q \in L^1(\mathcal{F}_T)$ occurring at time T. For simplicity, we restrict ourselves to non-negative Q. Since γ is non-constant, we cannot derive closed-form formulas for the bid and ask price as in the constant liquidity case. We therefore have to calculate them backward recursively according to Definition 9. To that end, we define $b^{\Psi,\Gamma}_{t_K}(Q) = a^{\Psi,\Gamma}_{t_K}(Q) = Q$ with $t_K = T$. Assuming we have already calculated $b^{\Psi,\Gamma}_{t_{k+1}}(Q)$ and $a^{\Psi,\Gamma}_{t_{k+1}}(Q)$, we can calculate bid and ask prices for every node of the previous time-step as follows. First, starting from a node (S_{t_k}, γ_{t_k}), we sort the four possible states ω_{uu}, ω_{ud}, ω_{du}, and ω_{dd} such that

$$b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_1) \geq b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_2) \geq b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_3) \geq b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_4) \tag{24}$$

and

$$a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_1) \geq a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_2) \geq a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_3) \geq a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_4), \tag{25}$$

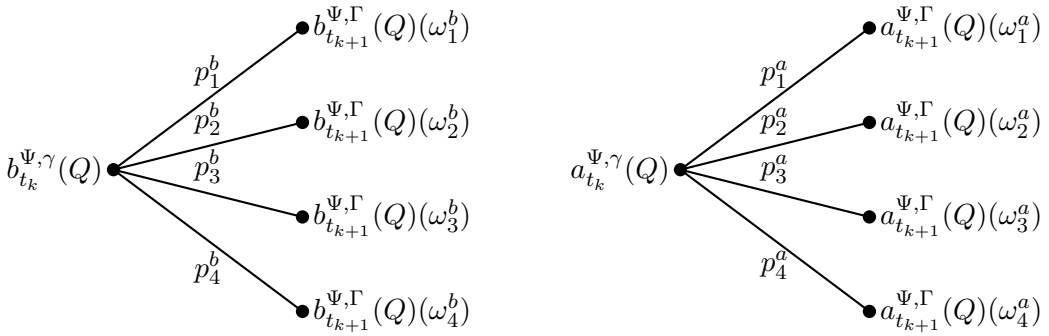

Figure 3. The same node but now the states and transition probabilities are sorted such that the bid resp. ask prices are highest for ω^b_1 resp. ω^a_1 and lowest for ω^b_4 resp. ω^a_4.

for states $\omega^b_i, \omega^a_i \in \{\omega_{uu}, \omega_{ud}, \omega_{du}, \omega_{dd}\}$, $i = 1, ..., 4$. In Figure 3, we plot the node for the bid and ask price. To simplify notation, we denote the transition probabilities corresponding to the states with the same sub- and superscripts. Then, since Q is non-negative, the bid price is given by

$$\begin{aligned} b^{\Psi,\Gamma}_{t_k}(Q) = \mathbb{E}^{\psi^{\gamma_{t_k}}}_{t_k}[b^{\Psi,\Gamma}_{t_{k+1}}(Q)] &= b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_1)(1-\psi^{\gamma_{t_k}}(p^b_2+p^b_3+p^b_4)) \\ &+ b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_2)(\psi^{\gamma_{t_k}}(p^b_2+p^b_3+p^b_4)-\psi^{\gamma_{t_k}}(p^b_3+p^b_4)) \\ &+ b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_3)(\psi^{\gamma_{t_k}}(p^b_3+p^b_4)-\psi^{\gamma_{t_k}}(p^b_4)) \\ &+ b^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^b_4)\psi^{\gamma_{t_k}}(p^b_4) \end{aligned} \tag{26}$$

and the ask price is

$$\begin{aligned} a^{\Psi,\Gamma}_{t_k}(Q) = -\mathbb{E}^{\psi^{\gamma_{t_k}}}_{t_k}[-a^{\Psi,\Gamma}_{t_{k+1}}(Q)] &= a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_1)\psi^{\gamma_{t_k}}(p^a_1) \\ &+ a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_2)(\psi^{\gamma_{t_k}}(p^a_1+p^a_2)-\psi^{\gamma_{t_k}}(p^a_1)) \\ &+ a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_3)(\psi^{\gamma_{t_k}}(p^a_1+p^a_2+p^a_3)-\psi^{\gamma_{t_k}}(p^a_1+p^a_2)) \\ &+ a^{\Psi,\Gamma}_{t_{k+1}}(Q)(\omega^a_4)(1-\psi^{\gamma_{t_k}}(p^a_1+p^a_2+p^a_3)). \end{aligned} \tag{27}$$

Continuing the backward recursion through the tree and applying the re-weighting of the reference probability measure according to the distortion function, we can calculate the bid and ask price of any non-negative future discounted cash flow at any point in time.

4 Application

To illustrate our methodology, we consider European options written on the S&P 500 index and we calibrate both a static and a stochastic liquidity model, as defined in Sections 3.3.1 and 3.3.2, to the bid-ask spreads of European index options.[15] All data comes from the OptionMetrics database accessed via the Wharton Research Data Services. For the calibration exercise, we choose as arbitrary date July 20, 2012. The S&P 500 had a closing price of $S_0 = 1362.66$ and the European options market was neither particularly stressed nor overly relaxed. At this day, the option data consists of 2560 calls and puts with maturities ranging from 7 days to 2.4 years and a strike interval of 100 to 3000. We removed 295 options from the dataset for whose mid price no Black-Scholes implied volatility could be calculated. For our discussion, we restrict our analysis to calls and puts within the $[80\%, 120\%]$ forward moneyness interval. Furthermore, we focus only on three maturities slices, namely on maturities of three, five, and eleven months. This leaves us with a total of 230 calls and puts.

4.1 Model specification

For our application, we assume that the log returns of the index are conditionally normal distributed.[16] For the liquidity process, we either assume it be constant for the static liquidity model or consider a mean-reverting square-root process following the findings of Albrecher, Guillaume, and Schoutens (2013) regarding the behavior of the static liquidity model parameter over time. In particular, for the stochastic liquidity model of Section 3.3.2, the asset price $S = (S_t)_{t \in \mathcal{T}_0^K}$ and the liquidity process $\Gamma = (\gamma_t)_{t \in \mathcal{T}_0^K}$ are binomial tree approximations of the continuous-time processes

$$dS_t = S_t(r-q)dt + S_t \sigma W_t^S \quad S_0 > 0, \tag{28}$$

$$d\gamma_t = \kappa(\theta - \gamma_t)dt + \nu\sqrt{\gamma_t} dW_t^\gamma, \quad \gamma_0 > 0, \tag{29}$$

[15] Using the same techniques, it is also possible to model bid and ask prices of path-dependent payoffs.

[16] This choice is merely for illustration purposes. We are well aware of the fact that there are more suitable choices for the underlying process. However, while, e.g., a stochastic volatility model such as the Heston model may be better suited, the imprecision in the volatility surface fit interferes with the assessment of the performance of the liquidity model.

where W^S and W^γ are correlated Brownian motions with $d\langle W^S, W^\gamma\rangle_t = \rho dt$, $\rho \in [-1, 1]$. By r we denote the one-period risk-free rate, q the dividend yield, and σ the asset volatility. To improve speed and memory usage, we prune the trees similar as in Baule and Wilkens (2004). For more details on the construction of the binomial trees, we refer to Appendix B. The asset price S in the static liquidity model is the same binomial tree approximation of (28) and the construction of the binomial tree follows similarly, except that γ is a constant now.

Following Madan, Pistorius, and Stadje (2015), we modify the usual concave distortion function ψ^γ to take into account the step size h of the tree via

$$\psi^{\gamma,h}(u) := u + \sqrt{h}(\psi^\gamma(u) - u), \quad \forall u \in [0, 1], \tag{30}$$

which is still concave. Using this adjustment allows for an adequate comparison of parameter estimates for the market liquidity process for different step sizes and in particular the one-step static liquidity model. As distortion measure, we use CVaR.[17]

4.2 Calibration

In the following, we only refer to the calibration of the stochastic liquidity model. The calibration of the static model follows the same methodology and only differs in the constructed binomial tree (one-dimensional instead of two-dimensional) and number of liquidity parameters (one instead of five).[18]

Considering the available data in the different slices, it is evident that they are not equally distributed over either moneyness nor maturity. For example, the three month slices have over twice as many data points as the either of the other two and the eleven months moneyness interval [110%, 112%] contains over 20% of all data points despite being only 5% of the whole interval. To ensure a better fit over the whole moneyness range and not just areas with clustered strikes, we therefore calculate regularly interpolated bid and ask prices, P_i^{bid} and P_i^{ask} for $i \in \{1, ..., M\}$, for each slice.

[17] We also conducted our analysis by using different measures. The differences were insignificant.

[18] We remark that for our calibration exercise we follow the standard practice of calibrating, e.g., a stochastic volatility model, i.e., we treat the current level of liquidity as an additional parameter. A time-series estimation of implied liquidity would require setting up, e.g., a suitable filter method, which is beyond the scope of this paper.

These interpolated prices, instead of the real ones, will be used in our optimization algorithm below. The reported model fit in Table 1 and all figures will be based on real data.

We calibrate our model for each selected maturity slice of calls and puts separately. First, we calculate for each option $i \in \{1, ..., M\}$ the Black-Scholes implied volatility σ_i such that

$$\frac{1}{2}(P_i^{\text{ask}} + P_i^{\text{bid}}) = BS(S_0, K_i, T_i, r_i, q_i, \sigma_i),$$

where K_i and T_i are the strike and time-to-maturity of option i, respectively, r_i is the zero rate with maturity closest to T_i and q_i is the implied dividend rate from the put-call parity.[19] Furthermore, we denote the market bid and ask prices by P_i^{ask} and P_i^{bid}. For every option, the implied volatility parameter is then used to construct the binomial tree for the underlying together with the correlated tree for the liquidity process.[20] The risk-free rate, dividend yield and the liquidity parameters, including the correlation parameter ϱ, are the same over all options of the selected maturity slice while the implied volatility is allowed to be different to best replicate the observed market data.

Having fixed the distortion functions $\Psi = (\psi^z)_{z \geq 0}$ and a market liquidity process Γ in (29), which depends on the set of parameters $\Theta = \{\gamma_0, \kappa, \theta, \nu, \rho\}$, we can calibrate the model by minimizing the root-mean-square error (RMSE) of the normalized bid-ask spreads:[21]

$$\text{RMSE}(\Theta \,|\, \Psi, \Gamma) := \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(\frac{\Delta P_i^{\text{model}} - \Delta P_i^{\text{market}}}{\frac{1}{2}(P_i^{\text{bid}} + P_i^{\text{ask}})} \right)^2}, \tag{31}$$

where $\Delta P_i^{\text{market}}$ and $\Delta P_i^{\text{model}}$ are the market and model bid-ask spreads and M is the number of options. The model bid and ask prices are calculated by starting at the terminal cash flows and applying equations (26) respectively (27) backward recursively throughout the two-dimensional

[19] Inferring the dividend yield from the put-call parity partially circumvents problems caused by different quotation times from the option and underlying markets. The implied dividend yield for all options lies in the interval [1.90%, 2.21%]. The dividend yield reported in the OptionMetrics database is 2.50%.

[20] We remark that this procedure is only an approximation as the undistorted price is usually not exactly half-way between the bid and ask price in our model. The approximation becomes cruder when getting closer to maturity and further away from at the money. Nevertheless, compared to the alternative of having to calibrate M additional parameters, the error from this approach seems acceptable.

[21] We also tested minimizing the spreads in implied volatility, but found that the improved fit in the implied volatility space led to a, comparatively, bigger error in the normalized price space. Minimizing the model errors of the bid and ask prices respectively implied volatilities instead of the spreads did not lead to vastly different results.

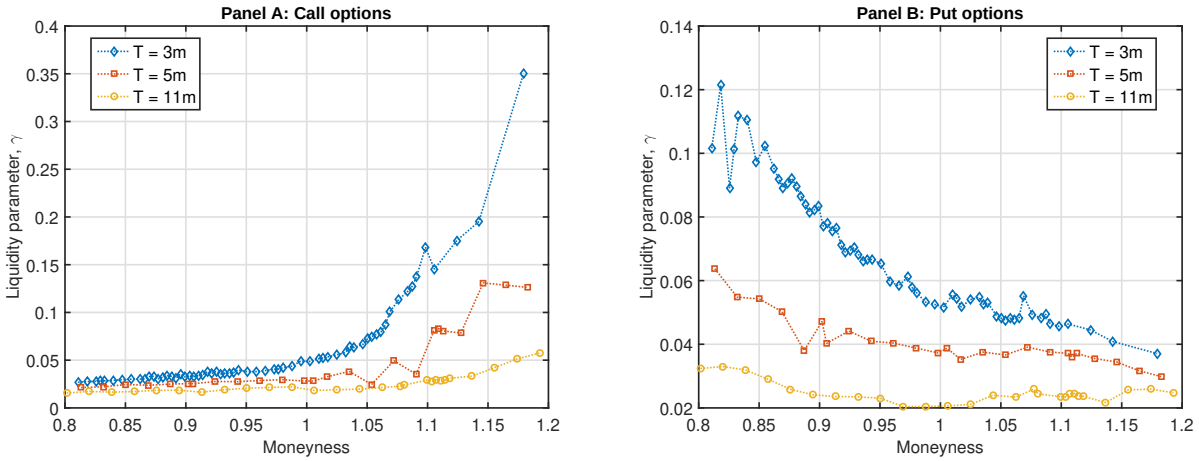

Figure 4. Market implied liquidity skew for selected maturity slices of the European options market on the S&P 500 on July 20, 2012. In total, we have 115 calls and the same number of puts. For every option, we calculate the static market liquidity parameter γ, which minimizes the bid-ask spread. We use the CVaR distortion function and assume a lognormal model for the underlying index.

binomial tree to get

$$\Delta P_i^{\text{model}} := a_0^{\Psi,\Gamma}(Q_i) - b_0^{\Psi,\Gamma}(Q_i), \tag{32}$$

where Q_i is the cash flow associated to the i^{th} option.[22]

To avoid the problem of local minima, we use a surrogate model based on radial basis functions.[23] Surrogate models are widely used in engineering because they require significantly less function evaluations than, e.g., genetic algorithms or particle swarm methods.

We first determined reasonable parameter ranges and optimized over these.[24] To help the numerical algorithm find better solutions, we then tightened the parameter ranges by determining where the best and worst model fits occurred. We tried to keep the parameter ranges as wide as possible to prevent influencing the final results too much.

4.3 Results and discussion

To motivate the use of a stochastic liquidity component, we first consider a static liquidity model as in Section 3.3.1 for which we regard every option in isolation. For the CVaR distortion function, we calculate for every option $i \in \{1, ..., M\}$ the unique parameter γ_i such that the model bid-ask spreads calculated using Corollary 1 coincide with the market bid-ask spreads. In Figure 4, we plot the calibrated liquidity parameter γ for various maturity slices and levels of moneyness. Clearly, this market-implied liquidity parameter is far from being constant. It increases with decreasing maturity and when the option becomes out-of-the-money. Furthermore, the skew effect weakens with increasing maturity.[25]

Going one step forward, we calibrate our stochastic liquidity model given in equation (29). As in the static model, we calibrate each maturity slice separately. Table 1 reports the parameter estimates and the RMSEs for the calibration of three maturity slices. As in the static case, the current implied liquidity level γ_0 decreases with increasing maturity. Furthermore, liquidity shocks tend to be highly transitory, which is reflected by the high values for the estimates of the parameter κ. These values fit well our observation of the bid-ask spreads over time in Panel B of Figure 1. The persistence of liquidity shocks, however, tends to increase with increasing maturity. The volatility estimate ν decreases with increasing maturity. We find that for call options the long-term mean θ is almost at the same level for all three maturities. For puts, we find a similar pattern, except for the eleven months maturity slice. Finally, the estimates for the correlation parameter ρ indicate a consistent and rather strongly negative correlation between changes in price and liquidity. Such high values make intuitively sense, as in times of highly negative returns such as, e.g., during the global financial crisis in 2008-2009 and the European crisis in 2012, bid-ask spreads also widened significantly.[26]

For comparison, we have also calibrated the static liquidity model for which all options on the

[22]For the static liquidity model the relevant formulas are collected in Corollary 1.

[23]See Gutmann (2001) and in particular the toolbox MATSuMoTo developed by Mueller (2014).

[24]$\gamma \in [0.5\gamma^*, 2\gamma^*]$, $\kappa \in (0, 100]$, $\theta \in (0, 0.05]$, $\nu \in (0, 2]$, $\varrho \in (-1, 1)$. γ^* denotes the optimal liquidity parameter in the static model, see the next section for more details.

[25]These properties of the implied liquidity parameter γ is in line with what has been observed in the static one-step models used in Albrecher, Guillaume, and Schoutens (2013) and Corcuera, Guillaume, Madan, and Schoutens (2012).

[26]The high correlation value corroborates the findings in Albrecher, Guillaume, and Schoutens (2013) of a high static liquidity parameter during times of crisis.

		Stochastic liquidity model						Static model	
Type	Mat	γ_0	κ	θ	ν	ρ	RMSE	γ^*	RMSE
call	3m	0.151	22.823	0.008	0.980	-0.951	9.4%	0.170	20.1%
call	5m	0.068	11.278	0.012	0.719	-0.974	3.4%	0.074	10.1%
call	11m	0.031	18.903	0.011	0.553	-0.983	1.4%	0.029	3.8%
put	3m	0.087	8.567	0.009	0.910	-0.975	1.2%	0.082	3.4%
put	5m	0.036	4.220	0.010	0.726	-0.952	0.7%	0.045	1.6%
put	11m	0.021	4.123	0.016	0.509	-0.977	0.6%	0.026	1.0%

Table 1. The table reports the calibrated parameters for European call and put options with different maturities (Mat) on the S&P 500 on July 20, 2012, for the stochastic liquidity model and the static model. The underlying distortion function is chosen to be CVaR. The parameters of the liquidity process are estimated by minimizing the RMSE of the normalized bid-ask spreads.

selected maturity slices are fitted simultaneously. The optimal static liquidity parameter γ^* turns out to be of similar magnitude than the estimated γ_0 in the stochastic liquidity model. However, the model fit is considerably worse, with an RMSE often more than twice as large as the RMSE from the stochastic model. This leads us to the conclusion that the stochastic liquidity model can lead to significant improvements statistically, while also representing more closely stylized facts about market liquidity.[27]

To further illustrate our model's capability of fitting bid-ask spreads, we plot in Figure 5 the bid-ask prices of calls and puts in terms of implied volatilities. We first observe that the general behavior of the data is well replicated by the model for all slices and both option types. The model performs worse for shorter maturities, where spreads are also generally larger than for longer maturities. However, especially for short-term OTM put options, the bid-ask spread is fitted remarkably well. In contrast, as can be observed in Figure 6, the static model struggles to replicate the bid-ask spread at short maturities. Especially the errors for ITM options tend to be substantial.

[27] As an additional exercise, we have re-calibrated our model on October 10, 2008, in the wake of the financial crisis. Most parameters were all of the same order of magnitude. The calibration yielded parameters θ and ν that were higher than on July 20, 2012. Furthermore, the current level of liquidity γ_0 was also larger. We attribute this observation to the increased uncertainty in the market at the time and the much wider bid-ask spreads, especially for puts. The observations that κ is much higher for calls than puts and that ν decreases with maturity also held on October 10, 2008. Comparing with the static liquidity model, the RMSE was again improved by a factor of about two to three.

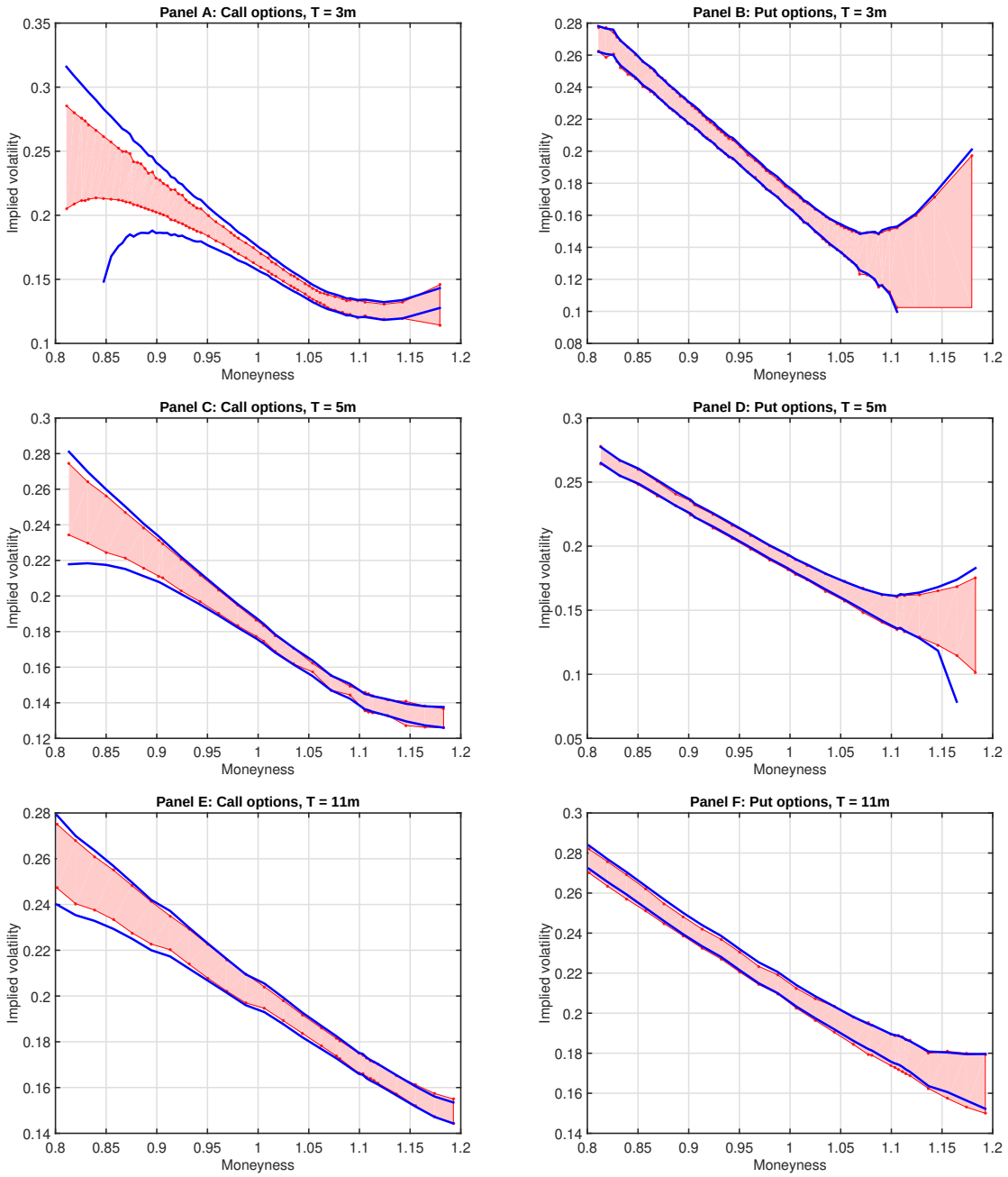

Figure 5. Calibration fit in the implied volatility space for European calls and puts on the S&P 500 on July 20, 2012. We plot the bid-ask spreads for maturities of 3, 5, and 11 months. We use the CVaR distortion function. The parameters of the liquidity process are estimated by minimizing the RMSE of the normalized bid-ask spreads. The shaded area corresponds to the market bid-ask spreads. The solid line corresponds to the model-implied spread.

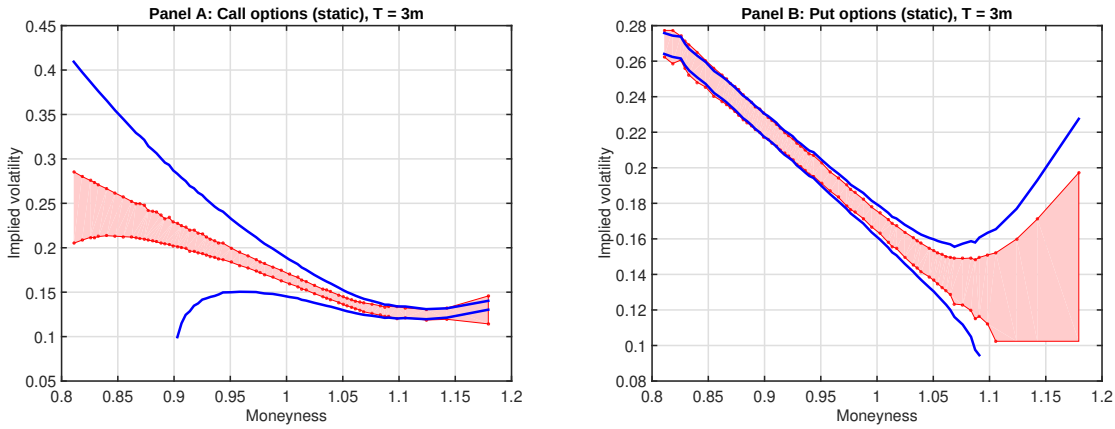

Figure 6. Calibration fit for the European call and put three months maturity slice on the S&P 500 on July 20, 2012. Panel A plots the bid-ask spread for calls, Panel B plots the spread in the implied volatility space for puts. The distortion function is chosen to be CVaR. The shaded area corresponds to the market bid-ask spreads. The solid line corresponds to the model-implied spread.

4.4 Parameter sensitivities

To gain more insights into the different roles of the parameters determining the liquidity process, we perform a sensitivity analysis for the put option with 5 months to maturity. Using the parameter values from Table 1, we assume different values for one parameter while keeping all others fixed. Figures 7 and 8 illustrate the sensitivities in the implied volatility and the normalized price space. Market implied bid-ask volatilities and prices correspond to the dashed lines marked with asterisks and are interpolated from S&P500 data on July 20, 2012.

Panels A and B of Figure 7 plot the impact of changes in the current liquidity level γ_0. As expected, high illiquidity, i.e., a large value for γ_0, is reflected by a wider bid-ask spread. For low values of γ_0, the spread collapses. In the implied volatility space, we note that the level of liquidity seems to impact options across all levels of moneyness in a similar way, i.e., by a parallel move of the implied volatility curve. Panels C and D plot the impact of changes in the parameter κ. Finally, in Panels E and F we plot the impact of changes in the parameter θ. Both parameters relate to the drift of the liquidity process. As for γ_0 in Panel A, we observe that changes in these parameters lead to a parallel move in the implied volatility surface. For more persistent liquidity shocks, i.e., a high value for κ, the bid-ask spread widens and it further does so when the long term mean for illiquidity θ is large.

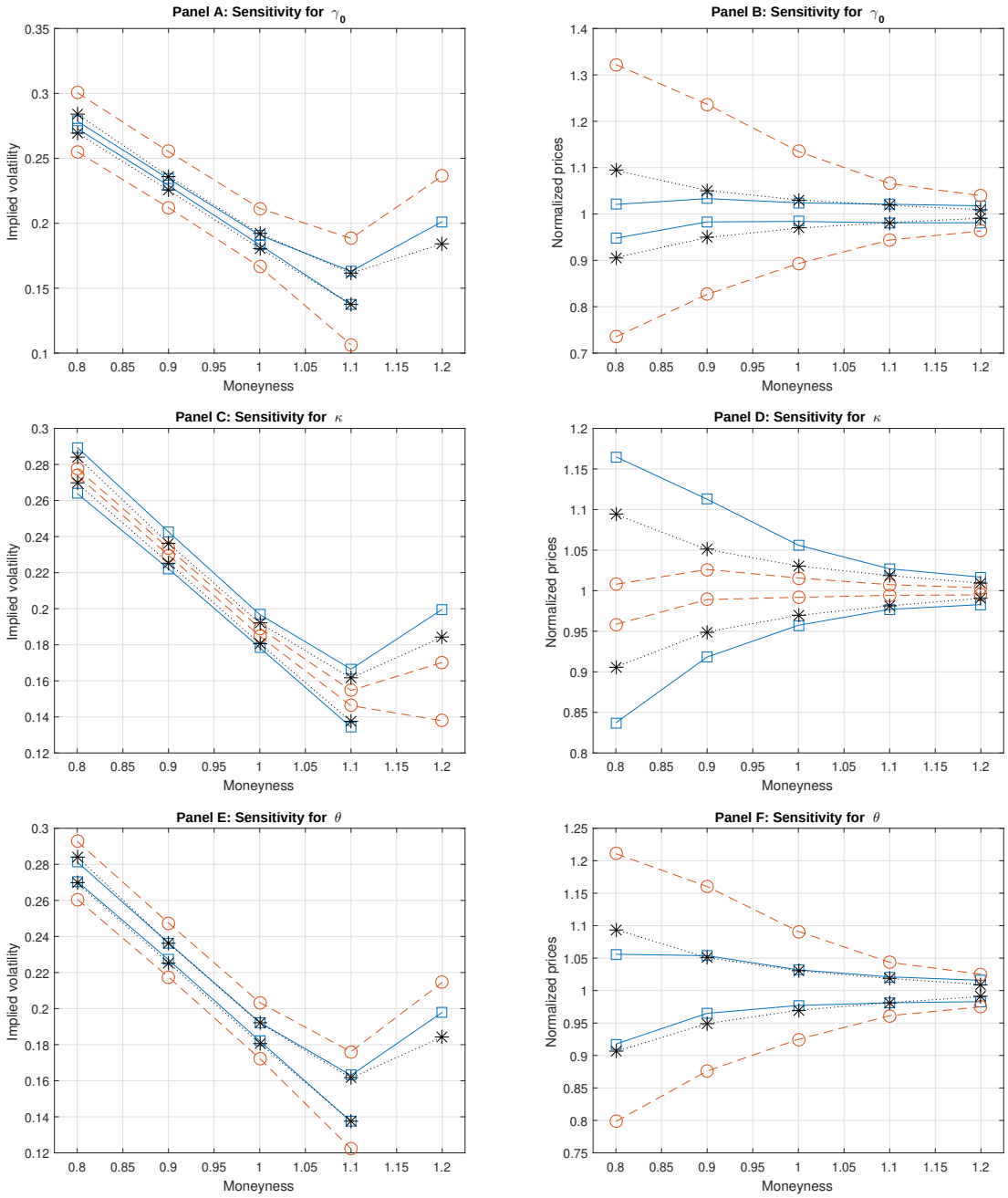

Figure 7. Sensitivity analysis for different parameter values. The figure plots the changes in the implied volatility and normalized price curve of the put option with 5 month maturity when we change the underlying liquidity parameters. Market implied bid-ask volatilities and prices correspond to the dashed lines marked with asterisks and are interpolated from S&P500 data on July 20, 2012. Panels A and B plot the impact of changes in γ_0. Solid (dashed) lines and squares (circles) correspond to $\gamma_0 = 0.01$ ($\gamma_0 = 0.2$). Panels C and D plot the impact of changes in the parameter κ. Solid (dashed) lines marked with squares (circles) correspond to $\kappa = 1$ ($\kappa = 30$). Panels E and F plot the impact of changes in the parameter θ. Solid (dashed) lines and squares (circles) correspond to $\theta = 0.001$ ($\theta = 0.1$). For all graphs, the remaining parameters were set equal to the values given in Table 1.

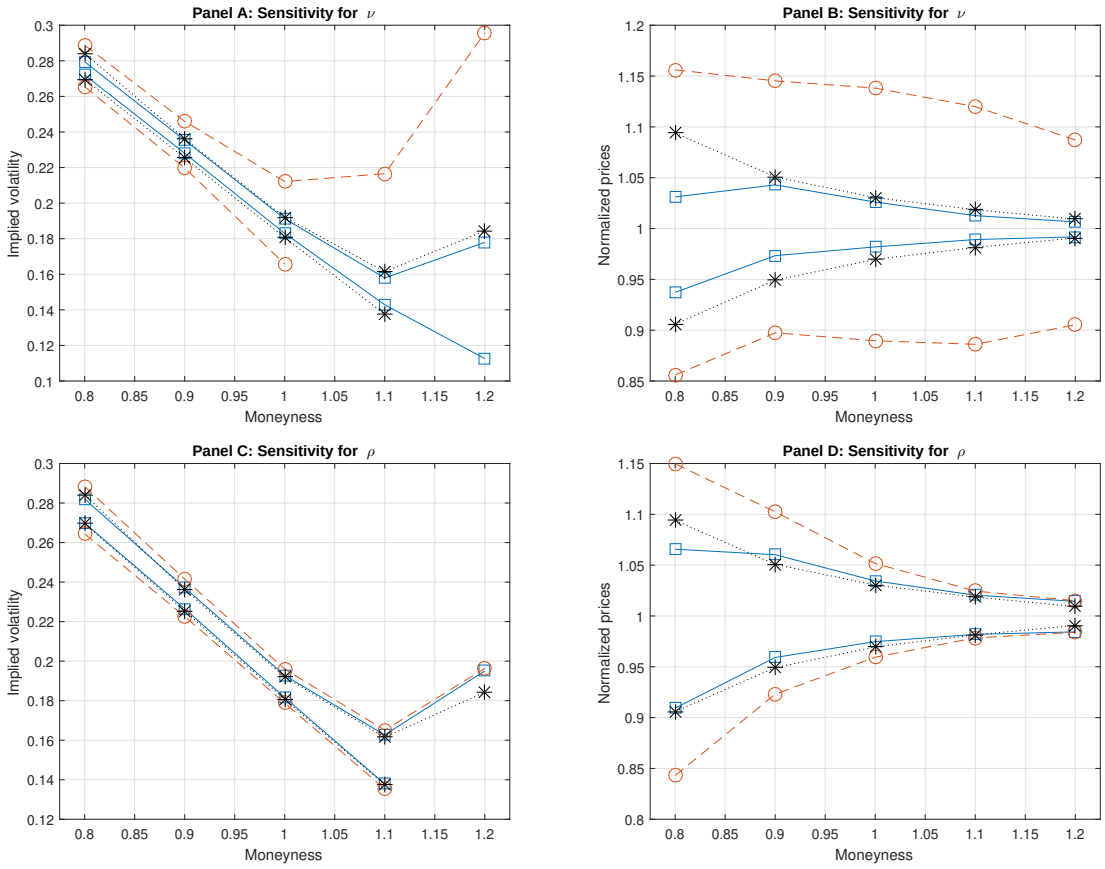

Figure 8. Sensitivity analysis for different parameter values. The figure plots the changes in the implied volatility and normalized price curve of the put option with 5 month maturity when we change the underlying liquidity parameters. Market implied bid-ask volatilities and prices correspond to the dashed lines marked with asterisks and are interpolated from S&P500 data on July 20, 2012. Panels A and B plot the impact of changes in the parameter ν. Solid (dashed) lines marked with squares (circles) correspond to $\nu = 0.25$ ($\nu = 0.15$). Panels C and D plot the impact of changes in the parameter ρ. Solid (dashed) lines and squares (circles) correspond to $\rho = -0.9$ ($\rho = 0.9$). For all graphs, the remaining parameters were set equal to the values given in Table 1.

In Figure 8 we plot the sensitivities with respect to the volatility of the liquidity process ν and the correlation between price changes and liquidity ρ. Here, we observe that the effect differs from the previous analysis in Figure 8 in that there is a less pronounced parallel impact on implied volatility. Instead, a change in the liquidity volatility parameter ν leads to a sharp increase in the convexity of the implied volatility of ask prices. Such an effect, but to a lesser extent, can also be observed for the implied correlation ρ.

5 Conclusion

Particularly after the financial crisis, the issue of market liquidity has been center stage for researchers and practitioners alike. However, the literature on liquidity in option markets and the discussion on how to incorporate liquidity into the pricing problem has been sparse. We add to this literature by providing a theoretical framework that allows us to incorporate a stochastic liquidity process into the option pricing problem. By specifying a simple version of our model and taking it to the data, we find that the additional flexibility of having a stochastic liquidity component helps to replicate the bid and ask spreads typically observed in option markets.

In our empirical exercise, we focus on a very simple and illustrative example. Hence, it comes at no surprise that our calibration analysis indicates some challenging avenues for future research. For example, the lack of fit at short-maturities corroborates the need for adding a jump component to the liquidity process. Furthermore, it may be advantageous to model the underlying process using a stochastic volatility model which, of course, leads to challenging numerical problems. A natural candidate for such an extension could be the recombining stochastic volatility tree of Akyıldırım, Dolinsky, and Soner (2014) as starting point. Inspired by the findings of Chou, Chung, Hsiao, and Wang (2011) it could be beneficial to consider a two-factor liquidity model, where one factor corresponds to the illiquidity of the underlying and the other to the illiquidity in the option market. Finally, one could go beyond a simple calibration exercise and try to perform a time-consistent estimation of the liquidity process using time series data. We leave these extensions to our analysis for future research.

References

Akyıldırım, Erdinç, Yan Dolinsky, and H Mete Soner, 2014, Approximating stochastic volatility by recombinant trees, *The Annals of Applied Probability* 24, 2176–2205.

Albrecher, Hansjoerg, Florence Guillaume, and Wim Schoutens, 2013, Implied liquidity: Model sensitivity, *Journal of Empirical Finance* 23, 48–67.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath, 1999, Coherent measures of risk, *Mathematical Finance* 9, 203–228.

——— , and Hyejin Ku, 2007, Coherent multiperiod risk adjusted values and Bellman's principle, *Annals of Operations Research* 152, 5–22.

Bannör, Karl F, and Matthias Scherer, 2014, On the calibration of distortion risk measures to bid-ask prices, *Quantitative Finance* 14, 1217–1228.

Barles, Guy, and Halil Mete Soner, 1998, Option pricing with transaction costs and a nonlinear Black-Scholes equation, *Finance and Stochastics* 2, 369–397.

Baule, Rainer, and Marco Wilkens, 2004, Lean trees - A general approach for improving performance of lattice models for option pricing, *Review of Derivatives Research* 7, 53–72.

Biagini, Sara, and Jocelyne Bion-Nadal, 2014, Dynamic quasi concave performance measures, *Journal of Mathematical Economics* 55, 143–153.

Bielecki, Tomasz R, Igor Cialenco, and Tao Chen, 2015, Dynamic conic finance via backward stochastic difference equations, *SIAM Journal on Financial Mathematics* 6, 1068–1122.

Bielecki, Tomasz R, Igor Cialenco, Ismail Iyigunler, and Rodrigo Rodriguez, 2013, Dynamic conic finance: Pricing and hedging in market models with transaction costs via dynamic coherent acceptability indices, *International Journal of Theoretical and Applied Finance* 16, 1350002.

Bongaerts, Dion, Frank De Jong, and Joost Driessen, 2011, Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market, *The Journal of Finance* 66, 203–240.

Çetin, Umut, Robert Jarrow, Philip Protter, and Mitch Warachka, 2006, Pricing options in an extended Black Scholes economy with illiquidity: Theory and empirical evidence, *Review of Financial Studies* 19, 493–529.

Çetin, Umut, Robert A Jarrow, and Philip Protter, 2004, Liquidity risk and arbitrage pricing theory, *Finance and Stochastics* 8, 311–341.

Chan, Kalok, and Y. Peter Chung, 2012, Asymmetric price distribution and bid–ask quotes in the stock options market, *Asia-Pacific Journal of Financial Studies* 41, 87–102.

Cherny, Alexander, and Dilip Madan, 2009, New measures for performance evaluation, *Review of Financial Studies* 22, 2571–2606.

Choquet, Gustave, 1953, Theory of capacities, in *Annales de l'Institut Fourier* vol. 5 pp. 131–295.

Chou, Robin K, San-Lin Chung, Yu-Jen Hsiao, and Yaw-Huei Wang, 2011, The impact of liquidity on option prices, *Journal of Futures Markets* 31, 1116–1141.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2015, Illiquidity premia in the equity options market, *Working Paper*.

Cohen, Samuel N, and Robert J Elliott, 2010a, Comparisons for backward stochastic differential equations on Markov chains and related no-arbitrage conditions, *The Annals of Applied Probability* 20, 267–311.

——— , 2010b, A general theory of finite state backward stochastic difference equations, *Stochastic Processes and their Applications* 120, 442–466.

——— , 2011, Backward stochastic difference equations and nearly time-consistent nonlinear expectations, *SIAM Journal on Control and Optimization* 49, 125–139.

Coquet, François, Ying Hu, Jean Mémin, and Shige Peng, 2002, Filtration-consistent nonlinear expectations and related g-expectations, *Probability Theory and Related Fields* 123, 1–27.

Corcuera, Jose Manuel, Florence Guillaume, Dilip B Madan, and Wim Schoutens, 2012, Implied liquidity: towards stochastic liquidity modelling and liquidity trading, *International Journal of Portfolio Analysis and Management* 1, 80–91.

Cvitanić, Jakša, and Ioannis Karatzas, 1996, Hedging and portfolio optimization under transaction costs: A martingale approach, *Mathematical Finance* 6, 133–165.

Davis, Mark HA, Vassilios G Panas, and Thaleia Zariphopoulou, 1993, European option pricing with transaction costs, *SIAM Journal on Control and Optimization* 31, 470–493.

Denneberg, Dieter, 1994, *Non-Additive Measure and Integral* (Springer).

El Karoui, Nicole, Shige Peng, and Marie Claire Quenez, 1997, Backward stochastic differential equations in finance, *Mathematical Finance* 7, 1–71.

Engle, Robert, and Breno Neri, 2010, The impact of hedging costs on the bid and ask spread in the options market, *Unpublished Working Paper, New York University.*

Ethier, Stewart N, and Thomas G Kurtz, 2009, *Markov Processes: Characterization and Convergence* (John Wiley & Sons).

Feng, Shih-Ping, Mao-Wei Hung, and Yaw-Huei Wang, 2014, Option pricing with stochastic liquidity risk: Theory and evidence, *Journal of Financial Markets* 18, 77–95.

Föllmer, Hans, and Alexander Schied, 2011, *Stochastic Finance: An Introduction in Discrete Time* (Walter de Gruyter).

George, Thomas J, and Francis A Longstaff, 1993, Bid-ask spreads and trading activity in the S&P 100 index options market, *Journal of Financial and Quantitative Analysis* 28, 381–397.

Gutmann, Hans-Martin, 2001, A radial basis function method for global optimization, *Journal of Global Optimization* 19, 201–227.

Madan, Dilip B, 2010, Conserving capital by adjusting deltas for gamma in the presence of skewness, *Journal of Risk and Financial Management* 3, 1–25.

——— , and Alexander Cherny, 2010, Markets as a counterparty: An introduction to conic finance, *International Journal of Theoretical and Applied Finance* 13, 1149–1177.

Madan, Dilip B, Martijn Pistorius, and Mitja Stadje, 2015, On dynamic spectral risk measures and a limit theorem, *arXiv preprint arXiv:1301.3531*.

Madan, Dilip B, and Wim Schoutens, 2014, Two processes for two prices, *International Journal of Theoretical and Applied Finance* 17, 1450005.

Mueller, Juliane, 2014, MATSuMoTo: The MATLAB surrogate model toolbox for computationally expensive black-box global optimization problems, *arXiv preprint arXiv:1404.4261*.

Pedersen, Lasse Heje, 2009, When everyone runs for the exit, *International Journal of Central Banking*.

Peng, Shige, 2007, G-expectation, G-Brownian motion and related stochastic calculus of Itô type, in *Stochastic Analysis and Applications*vol. 2 . pp. 541–567 (Springer).

Riedel, Frank, 2004, Dynamic coherent risk measures, *Stochastic Processes and their Applications* 112, 185–200.

Rosazza Gianin, Emanuela, and Carlo Sgarra, 2013, Acceptability indexes via g-expectations: an application to liquidity risk, *Mathematics and Financial Economics* 7, 457–475.

Shreve, Steven E, and H Mete Soner, 1994, Optimal investment and consumption with transaction costs, *The Annals of Applied Probability* 4, 609–692.

Soner, Halil M, Steven E Shreve, and J Cvitanić, 1995, There is no nontrivial hedging portfolio for option pricing with transaction costs, *The Annals of Applied Probability* 5, 327–355.

Wang, Shaun S, 2000, A class of distortion operators for pricing financial and insurance risks, *Journal of Risk and Insurance* 67, 15–36.

A Proof of Theorem 1

In the next two sections, we will summarize the theory of nonlinear expectations as solutions to BSΔEs on general discrete-time processes as developed by Cohen and Elliott (2010b). This framework is then used to show that the bid and ask prices from Definition 9 are in fact time-consistent and dynamically translation invariant nonlinear expectations as claimed in Theorem 1.

The log-price process of the underlying is denoted by $X = (X_t)_{t\in\mathcal{T}_0^K}$ which, like S, is a general discrete-time, finite state process. Cohen and Elliott (2010b) developed the theory for terminal conditions with multi-dimensional payoffs, but we will only present their results for the one-dimensional case. Additionally, while we assume without loss of generality that there are $N \in \mathbb{N}$ states at each time point $t \in \mathcal{T}_0^K$, it is worth noting that the theory can be extended to infinite states (see Cohen and Elliott (2011)), even though this is less relevant for our application which uses binomial trees.

A.1 BSΔE setup

Without loss of generality, we will set $X_0 = 0$ and assume that each X_{t_k}, $t_k \in \mathcal{T}_1^K$, takes values in the standard basis of $\mathbb{R}^N$, i.e.,

$$X_{t_k} \in \{e_1, \cdots, e_N\}, \quad e_j := (0, \cdots, 0, 1, 0, \cdots, 0)^\top \in \mathbb{R}^N,$$

where e_j is one at the j-th component and $(\cdot)^\top$ denotes the transposition operator. We call the process $M := (M_{t_k})_{k=0}^K$ defined by

$$M_{t_k} := X_{t_k} - \mathbb{E}[X_{t_k} \,|\, \mathcal{F}_{t_{k-1}}] \in \mathbb{R}^N, \quad k = 1, ..., K \tag{33}$$

and $M_0 := 0$ the martingale difference process.

Definition A.1 (BSΔE)**.** *Let $(Y, Z) := (Y_t, Z_t)_{t\in\mathcal{T}_0^K}$ be $\mathbb{R}\times\mathbb{R}^N$-valued adapted processes, $F : \Omega \times \mathcal{T}_0^K \times \mathbb{R} \times \mathbb{R}^N \longrightarrow \mathbb{R}$ an adapted function and Q an $\mathbb{R}$-valued, $\mathcal{F}_T$-measurable random variable.*

We say (Y, Z) is a solution of the BSΔE based on M with driver F and terminal condition Q if

and only if (Y, Z) *satisfies for all* $\omega \in \Omega$ *and* $t_k \in \mathcal{T}_0^K$,

$$Y_{t_k}(\omega) - \sum_{t_i \in \mathcal{T}_k^{K-1}} F(\omega, t_i, Y_{t_i}(\omega), Z_{t_i}(\omega)) + \sum_{t_i \in \mathcal{T}_k^{K-1}} Z_{t_i}(\omega) M_{t_{i+1}}(\omega) = Q(\omega). \tag{34}$$

From now on we will omit the argument $\omega \in \Omega$ of M, Q, X, Y and Z. Also note that the BSΔE can be equivalently written in difference form as

$$\begin{aligned} Y_{t_k} - F(\omega, t_k, Y_{t_k}, Z_{t_k}) + Z_{t_k} M_{t_{k+1}} = Y_{t_{k+1}} \quad \forall t_k \in \mathcal{T}_0^{K-1} \\ Y_T = Q. \end{aligned} \tag{35}$$

A BSΔE is the discrete-time version of a BSDE (see El Karoui, Peng, and Quenez (1997)), i.e., for all $t \in [0, T]$ and appropriately defined functions and variables

$$Y_t - \int_t^T F(\omega, u, Y_{u-}, Z_u) du + \int_t^T Z_u dM_u = Q. \tag{36}$$

Even though the solution to the BSΔE is given by the pair (Y, Z), we are not particularly interested in Z and will only implicitly use it. Indeed, subtracting from the BSΔE its conditional expectation shows that the process $(Z_{t_k} M_{t_{k+1}})_{t_k \in \mathcal{T}_0^{K-1}}$ can be expressed only in terms of $(Y_{t_{k+1}})_{t_k \in \mathcal{T}_0^{K-1}}$.

Lemma A.1. *If* (Y, Z) *are adapted solutions to the BSΔE, then it holds for* $t_k \in \mathcal{T}_0^{K-1}$

$$Z_{t_k} M_{t_{k+1}} = Y_{t_{k+1}} - \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]. \tag{37}$$

Proof. The BSΔE is given by, for $t_k \in \mathcal{T}_0^{K-1}$,

$$Y_{t_k} - F(\omega, t_k, Y_{t_k} Z_{t_k}) + Z_t M_{t_{k+1}} = Y_{t_{k+1}}. \tag{38}$$

Taking the conditional expectation on both sides gives

$$Y_{t_k} - F(\omega, t_k, Y_{t_k} Z_{t_k}) = \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}], \tag{39}$$

since Y, Z and F are adapted and $\mathbb{E}[M_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] = 0$. Subtracting these two equations gives the result. □

We will in the following assume that all solutions (Y, Z) satisfy

$$Y_t, Z_t \in L^1(\mathcal{F}_t) \quad \forall t \in \mathcal{T}_0^K, \tag{40}$$

that the terminal condition Q is in $L^1(\mathcal{F}_T)$ and that the driver fulfills

$$F(\omega, t, Y_t, Z_t) \in L^1(\mathcal{F}_t) \quad \forall \omega \in \Omega, t \in \mathcal{T}_0^K, Y_t, Z_t \in L^1(\mathcal{F}_t). \tag{41}$$

Definition A.2 (Equivalent processes)**.** *Let Z^1, Z^2 be two $\mathbb{R}^N$-valued adapted process. We call Z^1 and Z^2 equivalent at time $t_k \in \mathcal{T}_0^{K-1}$, denoted as $Z^1_{t_k} \sim_{M_{t_{k+1}}} Z^2_{t_k}$, if and only if*

$$Z^1_{t_k} M_{t_{k+1}} = Z^2_{t_k} M_{t_{k+1}} \quad \mathbb{P}\text{-a.s.} \tag{42}$$

Z^1 and Z^2 are equivalent and we write $Z^1 \sim_M Z^2$ if and only if Z^1 and Z^2 are equivalent at all times in $\mathcal{T}_0^{K-1}$.

Under suitable assumptions, it is now possible to show that each BSΔE has a solution that is unique up to equivalence.

Theorem A.1. *Assume the driver $F : \Omega \times \mathcal{T}_0^K \times \mathbb{R} \times \mathbb{R}^N \longrightarrow \mathbb{R}$ satisfies*

(i) for all $\mathbb{R}$-valued adapted processes Y, $\mathbb{R}^N$-valued adapted processes Z^1, Z^2, $t \in \mathcal{T}_0^K$ and $\mathbb{P}$-almost all $\omega \in \Omega$, if $Z^1 \sim_M Z^2$ then

$$F(\omega, t, Y_t, Z^1_t) = F(\omega, t, Y_t, Z^2_t). \tag{43}$$

(ii) for all $z \in \mathbb{R}^N$, $t \in \mathcal{T}_0^K$ and $\mathbb{P}$-almost all $\omega \in \Omega$, the map $y \longmapsto y - F(\omega, t, y, z)$ is a bijection from $\mathbb{R} \longrightarrow \mathbb{R}$.

Then, for all $Q \in L^1(\mathcal{F}_T)$, the BSΔE (34) has an adapted and $\mathbb{R} \times \mathbb{R}^N$-valued solution (Y, Z) that is unique up to indistinguishability for Y and $\sim_M$ for Z.

Proof. Cohen and Elliott (2010b), Theorem 2. □

Definition A.3 (Set of indices). *The $\mathcal{F}_{t_k}$-measurable set of indices of possible values of $X_{t_{k+1}}$ given $\mathcal{F}_{t_k}$ is defined as*

$$\mathbb{J}_{t_k} := \{j \in \{1, ..., N\} \,|\, \mathbb{P}(X_{t_{k+1}} = e_j \,|\, \mathcal{F}_{t_k}) > 0\}, \tag{44}$$

where the e_j are the standard basis of $\mathbb{R}^N$ as before and $t_k \in \mathcal{T}_0^{K-1}$.

A.2 Connection between BSΔEs and nonlinear expectations

Definition A.4 (Balanced driver). *A driver F is called balanced if and only if it satisfies assumptions (i) and (ii) of Theorem A.1 and furthermore, for all $Q^1, Q^2 \in L^1(\mathcal{F}_T)$, the corresponding BS$\Delta$E solutions (Y^1, Z^1), (Y^2, Z^2) satisfy*

(iii) for all $t_k \in \mathcal{T}_0^K$ and $\mathbb{P}$-almost all $\omega \in \Omega$,

$$F^1(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) - F^1(\omega, t_k, Y^2_{t_k}, Z^2_{t_k}) \geq \min_{j \in \mathbb{J}_{t_k}} \{(Z^1_{t_k} - Z^2_{t_k})(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\}. \tag{45}$$

and equality holds only if $Z^1_{t_k} \sim_{M_{t_{k+1}}} Z^2_{t_k}$.

(iv) if for all $t \in \mathcal{T}_0^K$ and $\mathbb{P}$-almost all $\omega \in \Omega$

$$Y^1_t - F(\omega, t, Y^1_t, Z^1_t) \geq Y^2_t - F(\omega, t, Y^2_t, Z^2_t), \tag{46}$$

then $Y^1_t \geq Y^2_t$ $\mathbb{P}$-a.s.

Definition A.5 (Normalized driver). *A driver F is called normalized if and only if for all $t \in \mathcal{T}_0^K$, $\mathbb{R}$-valued, $\mathcal{F}_t$-measurable processes Y and $\mathbb{P}$-almost all $\omega \in \Omega$, it holds that $F(\omega, t, Y, 0) = 0$.*

This brings us to the main theorem which gives a one-to-one connection between time-consistent nonlinear expectations and solutions to BSΔEs.

Theorem A.2. *Let $(\mathcal{E}(\cdot \,|\, \mathcal{F}_t))_{t \in \mathcal{T}_0^K}$ an time-consistent nonlinear expectation.*

Then, the following are equivalent.

(i) $(\mathcal{E}(\cdot \,|\, \mathcal{F}_t))_{t \in \mathcal{T}_0^K}$ is dynamically translation invariant.

(ii) There exists a driver F that is balanced, independent of Y and normalized such that for each $Q \in L^1(\mathcal{F}_T)$, $Y_t = \mathcal{E}(Q \,|\, \mathcal{F}_t)$ is a solution to the BSΔE (34) with terminal condition Q and driver F.

Furthermore, the two statements are connected via

$$F(\omega, t_k, Y_{t_k}, Z_{t_k}) = \mathcal{E}(Z_{t_k} M_{t_{k+1}} \,|\, \mathcal{F}_{t_k}) \quad \forall t_k \in \mathcal{T}_0^K. \tag{47}$$

Proof. Cohen and Elliott (2010b), Theorem 7. □

We will exploit this result by defining a driver F according to our one-period intuition that satisfies the assumptions of Theorem A.2, (ii). Then, by taking the conditional expectation of the BSΔE (35), we get the backward recursive formula

$$\begin{aligned} \mathcal{E}(Q \,|\, \mathcal{F}_{t_k}) = Y_{t_k} &= \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] + F(\omega, t_k, Y_{t_k}, Z_{t_k}) \quad \forall t_k \in \mathcal{T}_0^{K-1} \\ Y_T &= Q, \end{aligned} \tag{48}$$

which allows us to calculate the time-consistent nonlinear expectation operator based on such a driver.

Remark A.1. *In the static model, coherent (distortion) risk measures are used to define nonlinear expectations. In this discrete-time extension it is the other way around as defining*

$$\varrho_t(Q) := -\mathcal{E}(Q \,|\, \mathcal{F}_t) \quad \forall t \in \mathcal{T}_0^K \tag{49}$$

gives a dynamic risk measure. This was proposed by Rosazza Gianin and Sgarra (2013) in a continuous-time Brownian Motion setting. Cohen and Elliott (2010a) prove that $(\varrho_t)_{t \in \mathcal{T}_0^K}$ satisfy all the necessary properties, provided the driver F fulfills the assumptions of Theorem A.2, (ii). In particular, translation invariance follows by normalization and independence of Y and monotonicity by F being balanced. Positive homogeneity requires F to be positively homogeneous itself. Furthermore, if F is convex, $(\varrho_t)_{t \in \mathcal{T}_0^K}$ is subadditive and hence coherent.

A.3 The bid and ask drivers

In general, $(\mathbb{E}_t^\psi[\cdot])_{t\in\mathcal{T}_0^K}$ is no time-consistent nonlinear and dynamically translation invariant expectation, though it still satisfies some of the properties as the following Lemma shows.

Lemma A.2. *Let $(\mathbb{E}_t^\psi[\cdot])_{t\in\mathcal{T}_0^K}$ be a distorted conditional expectation. Then, for all $t \in \mathcal{T}_0^K$, $\mathbb{E}_t^\psi[\cdot]$ satisfies for all $Q^1, Q^2 \in L^1(\mathcal{F}_T)$ and $q \in L^1(\mathcal{F}_t)$*

(i) adaptedness, i.e., $\mathbb{E}_t^\psi[q] = q$ $\mathbb{P}$-a.s.

(ii) monotonicity, i.e., if $Q^1 \leq Q^2$ $\mathbb{P}$-a.s. then $\mathbb{E}_t^\psi[Q^1] \leq \mathbb{E}_t^\psi[Q^2]$ $\mathbb{P}$-a.s. and in particular, equality holds if and only if $Q^1 = Q^2$ $\mathbb{P}$-a.s.

(iii) dynamic translation invariance, i.e., $\mathbb{E}_t^\psi[Q+q] = \mathbb{E}_t^\psi[Q] + q$ $\mathbb{P}$-a.s.

Proof. Let $t \in \mathcal{T}$.

(i) Let $q \in L^1(\mathcal{F}_t)$. Now, since q is $\mathcal{F}_t$-measurable, $\mathbb{1}_{\{q\leq x\}}$ is as well for all $x \in \mathbb{R}$ and hence,

$$\mathbb{P}(q \leq x \,|\, \mathcal{F}_t) = \mathbb{E}[\mathbb{1}_{\{q\leq x\}} \,|\, \mathcal{F}_t] = \mathbb{1}_{\{q\leq x\}} \quad \forall x \in \mathbb{R}. \tag{50}$$

This, combined with the fact that

$$\psi(\omega, \mathbb{1}_A) = \mathbb{1}_A \quad \forall \omega \in \Omega \tag{51}$$

for every set $A \in \mathcal{F}_T$, directly gives, for each $\omega \in \Omega$,

$$\begin{aligned}
\mathbb{E}_t^\psi[q](\omega) &= -\int_{-\infty}^0 \psi(\omega, \mathbb{1}_{\{q\leq x\}}(\omega))dx + \int_0^\infty 1 - \psi(\omega, \mathbb{1}_{\{q\leq x\}}(\omega))dx \\
&= -\int_{-\infty}^0 \mathbb{1}_{\{q\leq x\}}(\omega)dx + \int_0^\infty 1 - \mathbb{1}_{\{q\leq x\}}(\omega)dx \\
&= -\int_{-\infty}^0 \mathbb{1}_{\{q\leq x\}}(\omega)dx + \int_0^\infty \mathbb{1}_{\{q> x\}}(\omega)dx \\
&= q(\omega).
\end{aligned} \tag{52}$$

(ii) The monotonicity property of $\mathbb{E}_t^\psi[\cdot](\omega)$ for all $\omega \in \Omega$ follows by observing that the conditional probability measure in a single state, $P_t(\cdot)(\omega)$, is a probability measure. Therefore,

$\psi(\omega, P_t(\cdot)(\omega))$ is a finite monotone set function and hence the Choquet integral defined thereby is monotone.

(iii) Let $Q \in L^1(\mathcal{F}_T)$, $q \in L^1(\mathcal{F}_t)$, $x \in \mathbb{R}$ and $\omega \in \Omega$. Since q is $\mathcal{F}_t$-measurable, the conditional probability on $\mathcal{F}_t$ at ω of $\{Q + q \leq x\}$ only depends on $q(\omega)$, i.e.,

$$\mathbb{P}(Q + q \leq x \,|\, \mathcal{F}_t)(\omega) = \mathbb{P}(Q + q(\omega) \leq x \,|\, \mathcal{F}_t)(\omega). \tag{53}$$

We use this and the change of measure $y := x - q(\omega)$ to get

$$\begin{aligned}
\mathbb{E}_t^{\psi}[Q + q](\omega) &= -\int_{-\infty}^{0} \psi(\omega, \mathbb{P}(Q \leq x - q(\omega) \,|\, \mathcal{F}_t)(\omega))dx \\
&\quad + \int_{0}^{\infty} 1 - \psi(\omega, \mathbb{P}(Q \leq x - q(\omega) \,|\, \mathcal{F}_t)(\omega))dx \\
&= -\int_{-\infty}^{-q(\omega)} \psi(\omega, \mathbb{P}(Q \leq y \,|\, \mathcal{F}_t)(\omega))dy \\
&\quad + \int_{-q(\omega)}^{\infty} 1 - \psi(\omega, \mathbb{P}(Q \leq y \,|\, \mathcal{F}_t)(\omega))dy \\
&= \mathbb{E}_t^{\psi}[Q](\omega) + q(\omega).
\end{aligned} \tag{54}$$

□

Corollary A.1. *$(-\mathbb{E}_t^{\psi}[-\cdot])_{t \in \mathcal{T}_0^K}$ satisfies the same properties.*

Proof. Follows directly from Lemma A.2. □

We now define the drivers for the bid and ask price

Definition A.6 (Bid and ask driver)**.** *Let $\Psi = (\psi^z)_{z \geq 0}$ be a family of concave distortion functions that are pointwise increasing in z. For the market liquidity process $\Gamma = (\gamma_t)_{t \in \mathcal{T}_0^K}$, we define the bid driver to be the function given by*

$$F^{b,\Psi,\Gamma}(\omega, t_k, Y_{t_k}, Z_{t_k}) := \mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[Z_{t_k} M_{t_{k+1}}] \tag{55}$$

and the ask driver is set to be

$$F^{a,\Psi,\Gamma}(\omega, t_k, Y_{t_k}, Z_{t_k}) := -\mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[-Z_{t_k} M_{t_{k+1}}] \tag{56}$$

for all $\omega \in \Omega$, $t_k \in \mathcal{T}_0^K$, (Y, Z) $\mathbb{R} \times \mathbb{R}^N$*-valued adapted processes.*

By the definition of Choquet integrals and because we only consider concave distortion functions, the bid driver is convex and the ask driver is concave. Additionally, both are positively homogeneous and satisfy the following other properties.

Lemma A.3. *The drivers* $F^{b,\Psi,\Gamma}$ *and* $F^{a,\Psi,\Gamma}$ *satisfy the assumptions of Theorem A.2 (ii), i.e.,*

(i) independence of Y*.*

(ii) balanced on $L^1(\mathcal{F}_T)$*.*

(iii) normalization.

(iv) for each $Q \in L^1(\mathcal{F}_T)$*, there exists a solution to the BS*Δ*E (34) with terminal condition* Q *and driver* $F^{b,\Psi,\Gamma}$ *respectively* $F^{a,\Psi,\Gamma}$*.*

Proof. Denote $F^{b,\Psi,\Gamma}$ and $F^{a,\Psi,\Gamma}$ both by F.

(i) Holds by definition.

(ii) Due to the independence of Y and since $F(\omega, t_k, Y_{t_k}, Z^1_{t_k}) = F(\omega, t_k, Y_{t_k}, Z^2_{t_k})$ $\mathbb{P}$-a.s. if $Z^1_{t_k} \sim_{M_{t_{k+1}}} Z^2_{t_k}$ by definition, the only difficulty lies in proving that for all $\omega \in \Omega$ and $t_k \in \mathcal{T}_0^K$,

(a) for all $Q^1, Q^2 \in L^1(\mathcal{F}_T)$ and their corresponding solutions (Y^1, Z^1) and (Y^2, Z^2) it holds that $\mathbb{P}$-a.s.

$$F(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) - F(\omega, t_k, Y^2_{t_k}, Z^2_{t_k}) \geq \min_{j \in \mathbb{J}_{t_k}} \{(Z^1_{t_k} - Z^2_{t_k})(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}])\}. \tag{57}$$

(b) equality holds only if $Z^1_{t_k} \sim_{M_{t_{k+1}}} Z^2_{t_k}$.

The proof turns out to be essentially the same as the one of Theorem A.2 given by Cohen and Elliott (2010b), where the driver F satisfies the necessary properties of $(\mathcal{E}(\cdot \,|\, \mathcal{F}_{t_k}))_{t_k \in \mathcal{T}_0^K}$. Recall that

$$\mathbb{J}_{t_k} = \{i \in \{1, ..., N\} \,|\, \mathbb{P}(X_{t_{k+1}} = e_i \,|\, \mathcal{F}_{t_k}) > 0\}. \tag{58}$$

(a) Define a $\mathcal{F}_{t_k}$-measurable random variable q by

$$q := \min_{j \in \mathbb{J}_{t_k}} \{Y^1_{t_{k+1}} - Y^2_{t_{k+1}} \,|\, \mathcal{F}_{t_k}, X_{t_{k+1}} = e_j\}. \tag{59}$$

By Lemma A.1 we know that

$$Z_{t_k} M_{t_{k+1}} = Y_{t_{k+1}} - \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}], \tag{60}$$

where Y and Z denote Y^1 and Z^1 respectively Y^2 and Z^2. Plugging this into the definition of the random variable q gives

$$q = \mathbb{E}[Y^1_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] - \mathbb{E}[Y^2_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] + (Z^1_{t_k} - Z^2_{t_k}) \min_{j \in \mathbb{J}_{t_k}} \{e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\}, \tag{61}$$

where we used the definition of M as

$$M_{t_{k+1}} := X_{t_{k+1}} - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]. \tag{62}$$

By the definition of q, it holds that $\mathbb{P}$-a.s.

$$Y^1_{t_{k+1}} - q \geq Y^2_{t_{k+1}} \tag{63}$$

and since $\mathbb{E}^\psi$ is monotone by Lemma A.2, we have $\mathbb{P}$-a.s.

$$\mathbb{E}^\psi_{t_k}[Y^1_{t_{k+1}} - q] \geq \mathbb{E}^\psi_{t_k}[Y^2_{t_{k+1}}]. \tag{64}$$

However, since $\mathbb{E}^\psi$ is also dynamically translation invariant by the same Lemma, this is equivalent to

$$\mathbb{E}^\psi_{t_k}[Y^1_{t_{k+1}}] - \mathbb{E}^\psi_{t_k}[Y^2_{t_{k+1}}] \geq q. \tag{65}$$

Subtracting $\mathbb{E}[Y^1_{t_{k+1}} - Y^2_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]$ from both sides and using the definition of F and q as well as again the dynamic translation invariance of $\mathbb{E}^\psi$ gives the claim.

(b) We use the same random variable q as before and additionally assume equality which by

the previous argument is equivalent to assuming that

$$q = \mathbb{E}^{\psi}_{t_k}[Y^1_{t_{k+1}}] - \mathbb{E}^{\psi}_{t_k}[Y^2_{t_{k+1}}]. \tag{66}$$

Comparing this with expression (61) shows that

$$q = Y^1_{t_{k+1}} - Y^2_{t_{k+1}}. \tag{67}$$

Therefore,

$$Y^1_{t_{k+1}} - Y^2_{t_{k+1}} = \mathbb{E}^{\psi}_{t_k}[Y^1_{t_{k+1}}] - \mathbb{E}^{\psi}_{t_k}[Y^2_{t_{k+1}}] \tag{68}$$

which is equivalent to

$$Z^1_{t_k} \sim_{M_{t_{k+1}}} Z^2_{t_k}. \tag{69}$$

(iii) Holds by definition.

(iv) Holds by Theorem A.1.

□

Remark A.2. *In the following we will demonstrate that Madan's (2010) proof of showing that the drivers are balanced is incorrect. Let us rewrite Madan's argument using our notation.*

Proof in Madan (2010). Recall that,

$$M_{t_{k+1}} = X_{t_{k+1}} - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]. \tag{70}$$

Then, by the definition of the bid driver as the distorted conditional expectation of $Z_{t_k} M_{t_{k+1}}$ and because $X_{t_{k+1}}$ takes values in the standard basis of $\mathbb{R}^N$,

$$F^b(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) \geq \min_{j \in \mathbb{J}_{t_k}} \{Z^1_{t_k}(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\} \tag{71}$$

and

$$F^b(\omega, t_k, Y^2_{t_k}, Z^2_{t_k}) \leq \max_{j \in \mathbb{J}_{t_k}} \{Z^2_{t_k}(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\}. \tag{72}$$

Therefore,

$$F(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) - F(\omega, t_k, Y^2_{t_k}, Z^2_{t_k}) \geq \min_{j \in \mathbb{J}_{t_k}} \{(Z^1_{t_k} - Z^2_{t_k})(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\}. \tag{73}$$

□

The above line of reasoning is clearly wrong. What the argument used in the above proof of Madan (2010) actually shows is that

$$\begin{aligned} &F(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) - F(\omega, t_k, Y^2_{t_k}, Z^2_{t_k}) \\ &\qquad \geq \min_{j \in \mathbb{J}_{t_k}} \{Z^1_{t_k}(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\} + \min_{j \in \mathbb{J}_{t_k}} \{-Z^2_{t_k}(e_j - \mathbb{E}[X_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]\}. \end{aligned} \tag{74}$$

The minimizing indices j^*_1 of equation (71) and j^*_2 of equation (72) do not in general coincide with the minimizing index j^* of equation (73). Therefore, the lower bound of $F(\omega, t_k, Y^2_{t_k}, Z^1_{t_k}) - F(\omega, t_k, Y^2_{t_k}, Z^2_{t_k})$ that needs to hold for the balanced property, given in equation (73), is always greater or equal than the bound (74) achieved by the argument in Madan (2010). Hence, the above line of arguments does not prove that the bid driver is balanced. In contrast, our proof relies on the dynamic translation invariance of the driver with which the desired lower bound can be achieved. Furthermore, in the constant liquidity setting this property enables us to arrive at a, compared to Madan (2010), simplified formula for the bid and ask price.

A.4 The induced nonlinear expectation

Due to the dynamic translation invariance of the bid and ask drivers, we can arrive at a backward recursive formula for the nonlinear expectations induced by the bid and ask drivers defined in the previous section, only depending on Q, Ψ and Γ. This is a new result that was unobtainable before, since the dynamic translation invariance of the drivers was not taken into account.

Proof of Theorem 1. By Theorem A.2 and Lemma A.3, there exist time-consistent and dynamically translation invariant nonlinear expectations $\mathcal{E}^{a,\Psi,\Gamma}$ and $\mathcal{E}^{b,\Psi,\Gamma}$ that are solutions to the backward

recursion

$$\begin{aligned} \mathcal{E}(Q \,|\, \mathcal{F}_{t_k}) = Y_{t_k} &= \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] + F(\omega, t_k, Y_{t_k}, Z_{t_k}) \quad \forall t_k \in \mathcal{T}_0^{K-1} \\ Y_T &= Q, \end{aligned} \tag{75}$$

where F and $\mathcal{E}$ denote $F^{b,\Psi,\Gamma}$ and $\mathcal{E}^{b,\Psi,\Gamma}$ or $F^{a,\Psi,\Gamma}$ and $\mathcal{E}^{a,\Psi,\Gamma}$, respectively.

By Lemma A.1 we have

$$Z_{t_k} M_{t_{k+1}} = Y_{t_{k+1}} - \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}].$$

Additionally, $F^{a,\Psi,\Gamma}$ and $F^{b,\Psi,\Gamma}$ are dynamically translation invariant. Therefore, for all $t_k \in \mathcal{T}_0^{K-1}$,

$$\begin{aligned} Y_{t_k} &= \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] + F^{b,\Psi,\Gamma}(\omega, t_k, Y_{t_k}, Z_{t_k}) \\ &= \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}] + \mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[Y_{t_{k+1}} - \mathbb{E}[Y_{t_{k+1}} \,|\, \mathcal{F}_{t_k}]] \\ &= \mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[Y_{t_{k+1}}] \end{aligned} \tag{76}$$

and similarly for $F^{a,\Psi,\Gamma}$. By starting at $Y_T = Q$ and applying backward recursion we arrive at

$$\mathcal{E}^{b,\Psi,\Gamma}(Q \,|\, \mathcal{F}_{t_k}) = \mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[\mathbb{E}_{t_{k+1}}^{\psi^{\gamma_{t_{k+1}}}}[\cdots \mathbb{E}_{t_{K-1}}^{\psi^{\gamma_{t_{K-1}}}}[Q]]] = b_{t_k}^{\Psi,\Gamma}(Q) \tag{77}$$

and

$$\mathcal{E}^{a,\Psi,\Gamma}(Q \,|\, \mathcal{F}_{t_k}) = -\mathbb{E}_{t_k}^{\psi^{\gamma_{t_k}}}[\mathbb{E}_{t_{k+1}}^{\psi^{\gamma_{t_{k+1}}}}[\cdots \mathbb{E}_{t_{K-1}}^{\psi^{\gamma_{t_{K-1}}}}[-Q]]] = a_{t_k}^{\Psi,\Gamma}(Q), \tag{78}$$

which concludes the proof. □

B Binomial tree construction

Let $T > 0$ denote maturity. We consider the continuous time processes

$$dS_t = S_t(r-q)dt + S_t \sigma W_t^S \quad S_0 > 0, \tag{79}$$

$$d\gamma_t = \kappa(\theta - \gamma_t)dt + \nu\sqrt{\gamma_t} dW_t^\gamma, \quad \gamma_0 > 0, \tag{80}$$

where W^S and W^γ are correlated Brownian motions with $d\langle W^S, W^\gamma\rangle_t = \rho dt, \rho \in [-1, 1]$. Here, σ denotes the implied volatility of a specific option's mid price. Using the transformation

$$\begin{aligned} x_t &:= \log S_t, && (81)\\ y_t &:= \frac{2}{\nu}\sqrt{\gamma_t} - \frac{\rho}{\sigma}x_t, && (82) \end{aligned}$$

we get

$$\begin{aligned} dx_t &= \mu_x dt + \sigma dW_t^S, && (83)\\ dy_t &= \mu_y(x_t, y_t)dt + dW_t^\gamma - \varrho dW_t^S, && (84) \end{aligned}$$

with

$$\begin{aligned} \mu_x &:= r - q - \frac{1}{2}\sigma^2, && (85)\\ \mu_y(x_t, y_t) &:= \left(\gamma_t(x_t, y_t)^{-\frac{1}{2}}\left(\frac{\kappa}{\nu}(\theta - \gamma_t(x_t, y_t))\right) - \frac{\nu}{4}\right) - \frac{\rho\mu_x}{\sigma}, && (86)\\ \gamma_t(x_t, y_t) &:= \left(\frac{\nu}{2}\left(y_t + \frac{\rho}{\sigma}x_t\right)\right)^2. && (87) \end{aligned}$$

Defining the independent Brownian motions

$$B^x \equiv W^S, \quad B^y \equiv \frac{1}{\sqrt{1-\rho^2}}(W^\gamma - \varrho B^1), \tag{88}$$

we can rewrite the dynamics of x and y as

$$\begin{aligned} dx_t &= \mu_x dt + \sigma dB_t^x, && (89)\\ dy_t &= \mu_y(x_t, y_t)dt + \sqrt{1-\varrho^2}dB_t^y. && (90) \end{aligned}$$

This decoupling allows us to have transition probabilities for x and y which only depend on whether x respectively y have previously moved up or down.

For a constant time step size $h := \frac{T}{K}$, where T is the maturity and K the number of time steps,

the processes x and y are discretized, for $k = 0, ..., K$, as random walks:

$$X_k^{(K)} := x_0 + \sqrt{h}\sigma \sum_{i=1}^{k} \xi_i^X, \tag{91}$$

$$Y_k^{(K)} := y_0 + \sqrt{h(1-\varrho^2)} \sum_{i=1}^{k} \xi_i^Y, \tag{92}$$

where $(\xi_i^X)_{i=1}^K$ and $(\xi_i^Y)_{i=1}^K$ are independent random variables with values in $\{\pm 1\}$. We use the convention that $X_0^{(K)} \equiv x_0$ and $Y_0^{(K)} \equiv y_0$. In addition, we define the filtration

$$\mathcal{F}_k := \sigma(\xi_1^X, \xi_1^Y, ..., \xi_k^X, \xi_k^Y),$$

for $k = 1, ..., K$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$, where Ω is the obvious state space. The probabilities $p_k := \mathbb{P}_{k-1}(\xi_k^X = 1)$ and $q_k := \mathbb{P}_{k-1}(\xi_k^Y = 1)$ are determined via moment matching, which guarantees the weak convergence of $(X^{(n)}, Y^{(n)})$ to (x, y) (see, e.g., Ethier and Kurtz (2009) and Akyıldırım, Dolinsky, and Soner (2014)). In particular, the first and second moments must satisfy

(i) $\mathbb{E}[X_k^{(K)} - X_{k-1}^{(K)} \,|\, \mathcal{F}_{k-1}] = \mu_x h + o(h)$,

(ii) $\mathbb{E}[Y_k^{(K)} - Y_{k-1}^{(K)} \,|\, \mathcal{F}_{k-1}] = \mu_y(X_{k-1}^{(K)}, Y_{k-1}^{(K)})h + o(h)$,

(iii) $\mathbb{E}[(X_k^{(K)} - X_{k-1}^{(K)})^2 \,|\, \mathcal{F}_{k-1}] = \sigma^2 h + o(h)$,

(iv) $\mathbb{E}[(Y_k^{(K)} - Y_{k-1}^{(K)})^2 \,|\, \mathcal{F}_{k-1}] = (1-\varrho^2)h + o(h)$.

The second moment conditions are fulfilled by construction. The first moment conditions lead to:

$$p_k = \mathbb{P}_{k-1}(\xi_k^X = 1) = \frac{1}{2}\left(1 + \frac{\mu_x\sqrt{h}}{\sigma}\right) \tag{93}$$

$$q_k = \mathbb{P}_{k-1}(\xi_k^Y = 1) = \frac{1}{2}\left(1 + \frac{\mu_y(X_{k-1}^{(K)}, Y_{k-1}^{(K)})\sqrt{h}}{\sqrt{1-\varrho^2}}\right). \tag{94}$$

As in, e.g., Akyıldırım, Dolinsky, and Soner (2014), we need to truncate the transition probabilities

such that they take values in the unit interval. Therefore, we use the modified definitions:

$$p_k = \max\left\{0, \min\left\{1, \frac{1}{2}\left(1 + \frac{\mu_x\sqrt{h}}{\sigma}\right)\right\}\right\}, \tag{95}$$

$$q_k = \max\left\{0, \min\left\{1, \frac{1}{2}\left(1 + \frac{\mu_y(X_{k-1}^{(K)}, Y_{k-1}^{(K)})\sqrt{h}}{\sqrt{1-\varrho^2}}\right)\right\}\right\}. \tag{96}$$